SPAWAR
Systems Center
San Diego

TECHNICAL REPORT 1896
March 2003

Knowledge Desk
Limited Objective Experiment (LOE)

H. M. Oonk
J. H. Schermerhorn
D. Glaser
Pacific Science and Engineering Group, Inc.

J. G. Morrison
SSC San Diego

Approved for public release;
distribution is unlimited.

SSC San Diego

TECHNICAL REPORT 1896
March 2003

Knowledge Desk
Limited Objective Experiment (LOE)

H. M. Oonk
J. H. Schermerhorn
D. Glaser
Pacific Science and Engineering Group, Inc.

J. G. Morrison
SSC San Diego

Approved for public release;
distribution is unlimited.

SPAWAR
Systems Center
San Diego

SSC San Diego
San Diego, CA 92152-5001

SSC SAN DIEGO
San Diego, California 92152-5001

T. V. Flynn, CAPT, USN
Commanding Officer

R. C. Kolb
Executive Director

ADMINISTRATIVE INFORMATION

The work described in this report was performed for the Collaborative Technologies Project Team (244210) of the Simulation and Human Systems Technology Division (244) of the Command and Control Department (240) of Space and Naval Warfare Systems Center, San Diego (SSC San Diego) with Pacific Science and Engineering Group, Inc., under contract number N66001-99-D-0050. Funding was provided by the Office of Naval Research (ONR), Cognitive and Neural Science Technology Division under program element 0602233N. The ONR program officer was Mr. Jerry Malecki.

Released by
Dr. Jeff Morrison
Principal Investigator

Under authority of
J. L. Martin, Head
Simulation and Human Systems
Technology Division

ACKNOWLEDGMENTS

This report was prepared as part of SSC San Diego's ongoing Command 21 project, sponsored by the Office of Naval Research, Cognitive and Neural Science Technology Division, with Gerald S. Malecki as Program Officer. Dr. Jeffrey G. Morrison, SSC San Diego, is the principal investigator on the Command 21 project. One goal of this project is the development and operational evaluation of multi-headed displays known as the Knowledge Desks for use in advanced military command centers ashore and at sea. This report describes an evaluation in the continuing efforts directed toward this goal.

The authors would like to thank the following individuals whose efforts and input contributed to this report:

- Mr. Tom Rossi and First Class Petty Officer Vernon Edwards, U.S. Naval War College.
- The students, instructors and staff at U.S. Naval War College who served as participants and Subject Matter Experts.
- Dr. Jeffrey Morrison, SSC San Diego
- Mr. Ronald Moore, Mr. Gene Averett, Dr. Richard Kelly, Dr. John Gwynne, Mr. David Butz, and Mr. Daniel Manes, Pacific Science and Engineering Group.
- Mr. Steve Francis, Sonalysts, Inc.

EXECUTIVE SUMMARY

Multiple monitor workstations are becoming more and more common in the military command and control environment due to the requirement to monitor and access large quantities of information while performing complex tasks and making complex decisions. The Office of Naval Research (ONR) sponsored Command 21 project addressed this requirement by developing a six-monitor display designed to facilitate information production and consumption by an individual user. Known as a Knowledge Desk (K-Desk), these displays were employed during several wargames as well as aboard ships to support command-level decision-making in operational command centers. Although the value of having additional monitors has been widely acknowledged, the question of *how many* monitors the warfighter really needs to support his/her various tasks (cognitive and otherwise) remains unanswered. To address that question, a Limited Objective Experiment (LOE) was conducted that assessed the relative costs and benefits of different display configurations from a performance standpoint. The results of the LOE provided recommendations that relate to requirements for future Fleet procurements and installations.

The LOE consisted of two experiments that were based on principle tasks performed by warfighters in operational command centers. In one experiment, participants assumed the role of an information producer and in the other as an information consumer. In the Producer experiment, participants were required to create an integrated "knowledge product" using information from many disparate sources. The tasks chosen for the producer were based on those performed by planner and analysis staff found in operational command centers. In the Consumer experiment, participants were required to monitor the status of an operational mission and maintain situation awareness as would be expected by a watchstander in an operational command center. In both experiments, participants were required to concurrently perform other tasks—communicating in chat sessions, responding to e-mails, monitoring a tactical display, etc. The amount of workload induced when performing these tasks was based on a survey of users in the Fleet. Performance on the various tasks was measured in various monitor conditions: one, two, three, four, and six monitors.

Performance in the experiments was primarily assessed in terms of the speed and accuracy with which the participants conducted their tasks (e.g., the timing of responses in a chat session, the timing and accuracy of responses to e-mail requests, etc.). Situation awareness was determined using the answers to questions embedded in e-mails and chats sent during the experiments. The accuracy and quality of the knowledge product created in the Producer task was rated by subject matter experts and additional trained raters. We analyzed performance on the various tasks separately and in aggregate.

As expected, the pattern of results was different across the various tasks performed by participants in both experiments. Overall, the four-monitor condition supported the best performance in both the Producer and Consumer experiments. Based on results of the two experiments, we make preliminary recommendations. First, four monitors are recommended for producer tasks that involve the simultaneous tasks of (1) creating products through the integration of multiple information sources, (2) monitoring incoming information, and (3) responding to requests for information. Second, at least four monitors are recommended for consumer tasks that involve the simultaneous tasks of (1) monitoring an operational situation, (2) monitoring of incoming information, and (3) responding to requests for information. Third, it is clear that the nature of and specific combinations of decision-making tasks clearly will have an impact on the optimum number of displays for a given workstation. Further research is needed to compare performance in these tasks in a more rigorous manner.

The results are discussed in terms of their context-dependency and viewed as a first step that will define the parameters under which more interesting issues related to multi-monitor workstations for

the warfighter can be explored. In particular, the layout of information used by the warfighter should be examined in order to determine, among other things: (1) the types of tasks that require multi-monitor displays, (2) the effects on cognitive workload, (3) the display configurations that best support cognitive processes involved in warfighter tasks, and (4) the effects of user control over display configuration on task performance.

CONTENTS

FIGURES

TABLES

INTRODUCTION

OBJECTIVE

Rapid advances in technology have made it possible for the warfighter to monitor multiple data sources at the same time. This is often done while simultaneously trying to create information products, such as status briefs, and performing a variety of other required tasks. One recently deployed product designed to help support the workload this induces is the Knowledge Desk (K-Desk). K-Desks are multi-monitor display systems composed of six 15-inch diagonal, flat-panel displays in a 2x3 configuration (see Figure 1). They are designed to facilitate information production and consumption by an individual user. K-Desks were used during the 2001 Global War Game and onboard USS *Carl Vinson* (CVN 70) during Operation Enduring Freedom. They were recently installed aboard USS *Constellation* (CV 64), and are slated for installation at a number of other ship- and shore-based sites. Results of evaluations of K-Desk usage in operational settings, as well as user comments, have shown that the K-Desks improve information integration, information production, and situation awareness (Oonk et al., 2002; Rogers et al., 2002). This report presents the results of a Limited Objective Experiment (LOE) that was conducted at the Naval War College (Newport, RI). The focus of the LOE was to compare participants' performance in typical warfighter tasks across various monitor conditions using a K-Desk.

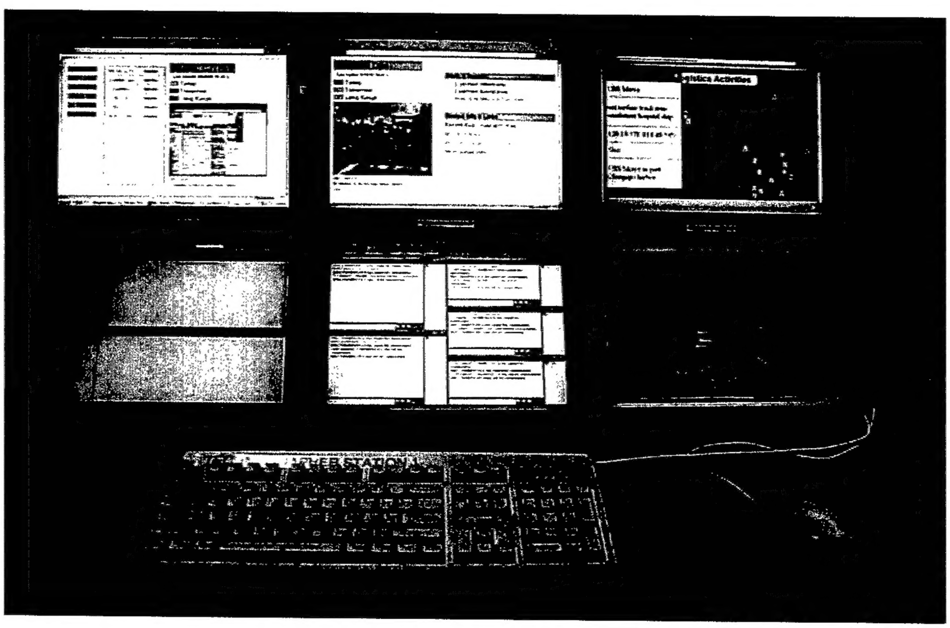

Figure 1. A K-Desk.

Although use of multiple monitors is becoming more common and the value of having additional monitors has been widely acknowledged anecdotally, few studies have attempted to evaluate multi-monitor workstations with the purpose of determining the optimum number of monitors. This is not surprising because the findings of such studies would be contingent on context—i.e., dependent on the number and nature of tasks. Research using multiple monitors has instead focused on more generalizable issues such as the ways in which people use and configure them (e.g., Grudin, 2001) or comparisons to other means of displaying multiple information sources (Card & Henderson, 1987; St. John, Harris, & Osga, 1997; St. John et al., 1999). The results of the experiments presented in this

report, likewise, are only useful if applied in settings where users are conducting the same or similar tasks.

With this in mind, we attempted to simulate realistic task environments based on feedback from potential users of the K-Desks (or other multi-monitor displays designed for the same purpose). A survey e-mailed to fleet users included questions about the amount of time spent conducting various tasks typical to the warfighter. They were also asked to report any tasks that were missing from the survey but should be included. Results of the survey were then incorporated in the design of the two experiments (see Appendix A for details of the survey and the results). Survey respondents fell into two basic categories:

- information producers – non-watchstanders who create information products to be used by others; and
- information consumers – watchstanders whose task is to monitor and use information created by others.

Likewise, the LOE consisted of two separate experiments: a Producer experiment, in which participants played the role of a Functional Component Commander (FCC), and a Consumer experiment in which they played the role of a Commander, Joint Task Force (CJTF).

BACKGROUND AND HYPOTHESES

Intuitively, there are many advantages to having multiple monitors when working on more than one task or document. First, they are a relatively inexpensive and flexible means to provide additional display real estate to a computer desktop. From a human factors perspective, multiple computer monitors reduce the need for interaction with the mouse and keyboard because they allow users to scan multiple information sources using only eye and head movements relative to a single monitor. Further, they reduce the need to "minimize" information (e.g., view one workspace or application while keeping others running "in the background"), which decreases users' reliance on working memory needed when switching between and/or integrating information across workspaces (Baddeley, 1986; St. John et al., 1999). The ability to view more than one workspace at a time may also prevent users from missing important changes or alerts that would otherwise occur in the "hidden" workspaces. Upon initial consideration, therefore, it seems that the more monitors, the better. However, our understanding of human factors and perception suggests that there are likely to be performance tradeoffs associated with increasing the number of available monitors.

Presenting multiple information sources simultaneously to users does not necessarily make it easier to integrate that information (e.g., Oonk et al., 2000). Too much information presented simultaneously may increase cognitive load (Sweller, 1988) associated with a cluttered visual environment. Further, increasing the amount of available screen space puts some information in the user's visual periphery, increasing the number and size of required eye, head, and mouse movements (Fitts, 1954; Gillan et al., 1990; Robinson, 1979; Whisenand & Emurian, 1999). In its "widest" configuration (i.e., at least three monitors are active[1]), information at the centers of the peripheral monitors of the K-Desk can be separated by 52° (60° separates information on the two farthest ends of the K-Desk). Previous research has suggested that looking at a target a small distance away from center (20°–30°) usually involves a single, discrete eye movement. However, viewing information that is more than 30° in the periphery requires additional eye and, sometimes, head movements (Robinson, 1979), each of which contributes additional motor programming and movement time.

[1] See the Method sections for each experiment for a description of the different monitor configurations examined in the two experiments.

Mouse movement times increase as the distance to the target increases, even for very short (4° or more) distances (Whisenand & Emurian, 1999). Placing information further away from the visual center also increases the detection times for even very salient visual events (Thackray & Touchstone, 1991), suggesting that users of multiple monitors may detect more slowly, or miss entirely, important alerts or changes that occur in peripheral monitors.

Several predictions about the results of the two experiments can be made based on the findings in literature and the tasks participants were required to perform. In both experiments, participants were required to visually search multiple information sources in order to gain and maintain an understanding of the operational situation. At the same time, they had to monitor messages that either provided them with new information or required them to answer questions about the situation (i.e., assessing their situation awareness)[2]. In the Producer task, participants had the additional role of creating an integrated information product based on information located in multiple sources. We hypothesized that:

- as the number of monitors increases beyond one, performance in all tasks would improve. While participants with fewer monitors would be required to flip back and forth between applications, participants with more monitors would be able to view more workspaces simultaneously. This should make it easier to integrate information from different sources and easier to monitor incoming messages, or alerts associated with them (e.g., because they did not appear in hidden workspaces); however,
- the increase in performance will reach a point of diminishing returns as the advantages to having multiple simultaneous views become accompanied by the costs of spreading information over a large area—such as additional and/or slower eye, head, and mouse movements;
- the point of diminishing returns will be reached with fewer monitors for the Producer experiment than the Consumer experiment. This is because participants in the Producer experiment will be required to make more mouse movements between monitors as they produce their integrated information product (e.g., when copying and pasting information between information sources).

[2] We use Endsley's (1995) definition of situation awareness as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."

PRODUCER EXPERIMENT

In the Producer LOE, participants assumed the role of a Functional Component Commander (FCC), reporting to a Commander, Joint Task Force (CJTF) in five fictional scenarios. Their primary task within each scenario was to create an integrated "information product," reporting what they believed to be the most important information for the CJTF. For each scenario, they were provided a collection of documents and pictures to browse using their K-Desk, in order to gain situation awareness. While browsing these documents to gain situation awareness, they were also required to monitor incoming information and answer questions using chat and e-mail. Their specific role and location was different for every scenario—over the course of the experiment, they played each of five FCCs[3] in each of five geographic locations.[4] Depending on the experimental condition, participants performed their tasks using one, two, three, four, or six monitors.[5]

Workload in these tasks was intended to simulate the workload of actual fleet users. Therefore, the number of inquiries that required responses, as well as the number of e-mails and chat rooms to monitor, was based on the results of a survey sent to users in the fleet prior to the experiment (see Appendix A). Performance in the experiment was based on the accuracy and quality the information product submitted to the CJTF and the speed and accuracy of chat and e-mail responses (which were used to assess participants' situation awareness).

METHOD

Participants

Thirty participants each served in a 2½ hour session. Between one and five participants performed in each session concurrently. Participants were instructors, students, or support personnel at the Naval War College. Demographic and computer experience information, collected at the conclusion of the experiment, is shown in Table 1. All participants were active or retired military (participants who reported rank as "civilian" were retired military), primarily Navy, with an average of 19.7 years of service.

---

[3] Functional Component Commanders were Intel, METOC, Logistics, Air Defense, and Force Protect.

[4] Scenarios took place in Cambodia, Korea, Bangladesh, China, and India.

[5] A five-monitor condition was not used for the following practical considerations: (1) investigators were concerned about participant fatigue during the experiment, as each monitor condition required 20 minutes of concentrated effort; and (2) a five-monitor display is not a likely configuration to be deployed due to physical awkwardness in installations and ergonomic considerations.

Table 1. Participant demographic information for the Producer and Consumer Experiments. [6]

		Producer Experiment (*n* = 29)	Consumer Experiment (*n* = 30)
Rank (Number of participants (% of total)	**O3**	2 (7%)	2 (7%)
	O4	4 (13%)	6 (20%)
	O5	14 (47%)	13 (43%)
	O6	5 (17%)	4 (13%)
	Civilian	2 (7%)	3 (10%)
	Not reported	2 (7%)	2 (7%)
Service (Years served in military)	**Mean**	19.7	19.9
	Standard Deviation	6.8	5.1
	Median	20.5	20.0
	Range	5-29	5-29
Computer/Software Experience[7] (% of participants with such experience)	**Web Browser**	100	100
	MS Chat	45	47
	MS NetMeeting	31	33
	MS Word	100	100
	MS PowerPoint	90	90
	MS Excel	83	87
	MS Outlook	100	100
	MS Log	31	27
	CommandNet	14	7
	Collaboration at Sea	0	0
	K-Web	14	13

Design

The design of the experiment was within-subject. Every participant served in each of the five display conditions (1, 2, 3, 4, and 6 monitors) presented in five 20-minute blocks. A Latin square was used to counterbalance the order of display conditions across participants. The five scenario-FCC combinations (Cambodia-Intel, Korea-MetOc, Bangladesh-Logistics, China-Air Defense, India-Force Protect) were presented in the same order for every participant.[8]

Procedure

Each participant performed all tasks using a K-Desk with 1, 2, 3, 4, or 6 windows activated. The software applications (e.g., *MS Chat*, *MS Outlook*) and electronic documents (a map and the scenario file folder) were displayed and configured on the desktop by the experimenter at the beginning of

---

[6] Note that no information was reported for one of the participants in the Producer Experiment because he did not complete the demographic survey administered at the end of the experiment.
[7] Chat®, NetMeeting®, Word®, Excel®, Outlook® are registered trademarks of the Microsoft Corporation.
[8] The order of scenarios was kept constant because multiple participants were run in the same experimental session. We did not believe this would be a problem because no comparisons were made across scenarios, and the order of the presentation for independent variable of interest—number of monitors—was counterbalanced.

each block (see Figure 2). Participants could move applications and documents into any active window of the K-desk as desired.

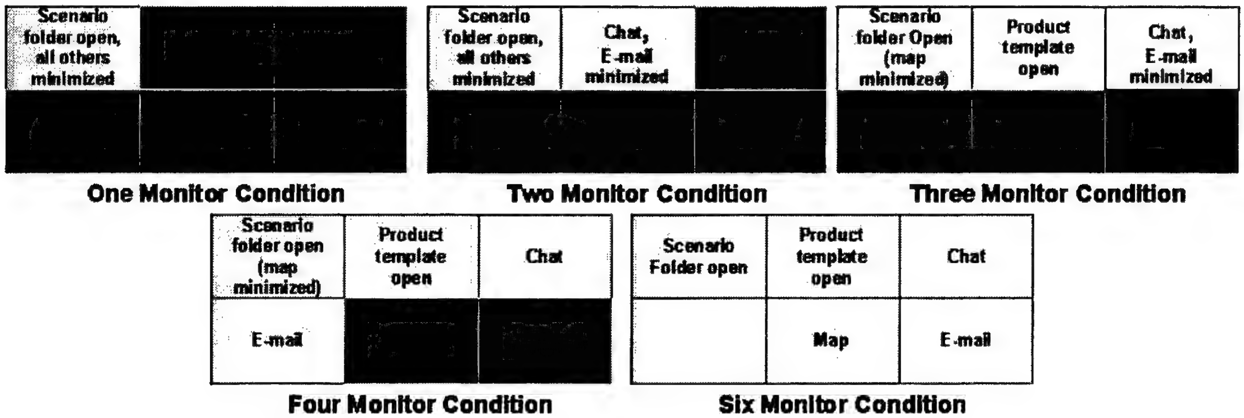

Figure 2. Initial display configurations for the monitor conditions in the Producer experiment. [9]

At the beginning of the experimental session, the participant received approximately 20 minutes of training. This included an overview of the task, followed by hands-on instruction on the following: (1) configuring the desktop, (2) switching between software applications, (3) using chat and e-mail software, and (4) filling out the information product template. Participants were given a chance to ask questions during and after the training session.

At the beginning of each experimental block, participants were read a brief description of the scenario and told which FCC role (e.g., Intel) they would be playing for that scenario. They were then told that they had 20 minutes to produce an information product to give to the CJTF. During this period, participants performed four tasks concurrently:

1. *Browsing scenario folders.* Participants were given access to an electronic folder (directory) containing files (average of 32 documents per scenario), which they could view or read to acquire situation awareness about the scenario. These files included *MS Word* documents (some with hyperlinks to other documents in the folder), web pages (some with hyperlinks to other documents in the folder), *MS PowerPoint* slides, and graphics (JPEG format). At least one of the documents in each scenario was a map.
2. *Creating the CJTF information product.* Participants were given a template (a simple *MS Word* table) to produce an information product for the CJTF. Participants were instructed to put as many pieces of information as they thought necessary into this document by (1) cutting and pasting from files in the scenario folders (either portions of, or entire documents), (2) cutting and pasting from chat or e-mail messages, or (3) typing directly into the template. Participants were told that these products would be rated on the extent to which they *included* all important and relevant information needed to provide situation awareness to the CJTF and *did not include* irrelevant content or excessive verbiage. They were told not to be concerned with format, layout, or "look and feel" of the product.

[9] We are aware that many monitor configurations could have been employed for each condition in these experiments. We were unfortunately restricted to these configurations because of the display settings of the K-Desks.

3. *Monitoring and responding to chat messages.* Participants were required to monitor two chat rooms and respond to any questions presented in them. During the course of the scenario, they received three chat messages: one or two of the messages contained new information about the scenario; the remainder contained a question posed by the CJTF. They were told that their performance would be rated on the speed and accuracy of their responses to these questions.

4. *Monitoring and responding to e-mail messages.* Participants were required to read and respond to e-mails. During the course of the scenario, they received five e-mail messages: one or two of the e-mails contained new information about the scenario; the remainder contained a question posed by the CJTF. They were told that their performance would be rated on the speed and accuracy of their responses to these questions.

In total, participants received three pieces of new information and five questions over the course of the scenario. The presentation of these was split between the chats and e-mails sent to the participants by confederates—the timing and text of these messages was determined by a script. Each experiment participant was assigned a confederate who played the role of CJTF and various other members of the command/battlegroup. The confederates were only allowed to send messages according to the script—they did not respond to spontaneous communications initiated by the participant[10]. The answers to the questions were available in the scenario folder documents or in earlier chats or e-mails.

At the end of each 20-minute block, participants e-mailed their information products to the CJTF. Participants were given a 5-minute break between blocks. At the end of the experiment, participants filled out a demographic information form and were asked to indicate how many monitors they thought best supported their tasks.

DATA ANALYSIS

Unless otherwise noted, the following information applies to the statistical analyses conducted. Statistical analyses were conducted using SPSS® 10.0. Descriptive statistics are furnished, including effect sizes and confidence intervals, when appropriate. Arcsines transformations were used to stabilize variances for proportions (Cohen et al., 2003). Given the nature of the design: response times, proportion correct, misses, adjusted scores, and web accesses were analyzed using repeated measures analysis of variance tests. All tests of significance used an alpha level of .05. Post hoc testing used the conservative Sidak test.

The following dependent variables were analyzed:

- *User preference.* At the end of the experiment, participants were asked to indicate what number of monitors best supported the experiment tasks.
- *Speed.* The response times for correct answers to the situation awareness questions presented and answered via e-mail and chat were analyzed for each experiment.
- *Accuracy.* Because of the nature of the questions being asked in the experiments, we believe that incorrect responses (i.e., questions that were answered, but incorrectly) should be treated separately from misses (i.e., questions that were never answered).[11]

---

[10] Communication in military command centers is often asynchronous, and therefore it was reasonable that some queries and comments might not receive a response for some period of time.

[11] A miss might occur for a number of reasons—e.g., if the participant had too many monitors, he may not have detected a new message because it had appeared in a monitor in his visual periphery; if he had too few monitors, he may have missed it because the chat or e-mail applications were minimized beneath another file or application.

Therefore, the following accuracy measures were employed for the chat and e-mail responses:

1. *Proportion Correct.* The proportion of *questions answered* that were correct responses.
2. *Misses.* The total number of questions for which there was no response.
3. *Adjusted Accuracy Score.* An adjusted proportion correct score that combines proportion correct and misses. Adjusted Accuracy Scores could range from 0.0 to 2.0 and were computed by the following formula (note that this 'penalizes' the participant more for a miss than an incorrect response):

   Adjusted Accuracy Score = (number correct responses + number responses) / (number misses + number responses)

- *Information product quality and accuracy.* Ratings of the information products were conducted by three subject matter experts (SMEs) at NWC and five additional trained raters (see Appendix B for the training document provided to all of the raters). The SMEs had an average of 21.33 years of military service (range = 18–24). The dependent variables that were derived from these ratings were:

  1. *A composite product score* for each product was based on ratings of individual items as "relevant," "irrelevant," or a mixture of relevant and irrelevant content. Based on these individual item scores, product accuracy was computed by the following formula. Each rater examined *all* items from all products of three different scenarios, and products from every scenario were rated by one SME and two other raters.

     Product Accuracy Score[12] = number items + ((number items + relevant items)/ (irrelevant items + 1))

     Every scenario was rated by three raters, one of which was a SME. Each rated all items (from all the products) of two different scenarios. The range of possible scores was 1.5 to 25.5.

  2. *The quality of each product* as a whole was rated on a scale from 1 to 3 in terms of:

     a. the extent to which it includes "right" information (1 = Includes none (or little) of the important/critical information, 2 = Includes a fair amount of the important/critical information, 3 = Includes most of the important/critical information);

     b. the extent to which it includes "wrong" information (1 = Includes mostly (or only) information that was *not* important/critical, 2 = Includes some information that *was not important/critical*, 3 = Includes no (or little) information that was *not* important/critical); and

     c. the extent to which the information is processed/filtered/interpreted for the CJTF (1 = Does *not* process (interpret or paraphrase) information or make connections, 2 = Does *some* processing of information and/or makes connections between item, 3 = Most or all of the content is processed. Connections are made between information items).

  Every scenario was rated by three raters, one of which was a SME, and each product was rated by at least two raters. Each rater examined at least half of the products of two different scenarios.[13]

---

[12] Items marked as a mixture of relevant and irrelevant were counted as .5 of a relevant item.

[13] Two of the SMEs and one of the trained raters only rated half of the information products due to time constraints.

RESULTS

Appendix C contains tables of the automated data presented in the figures in the Results sections.

Preference

As shown in Figure 3, most participants (n = 9) indicated that four monitors best supported their tasks. A single sample chi-square was conducted as a test of significance (or goodness of fit) for the preference data, given the categorical nature of the metric. This test allows us to examine if there are differential counts (or frequencies) for monitor preference, i.e., if there is a significant difference in the number of participants across the "monitor preference" categories. The analyses revealed a significant difference between the number of people who indicated best task support across monitor conditions ($\chi^2(6) = 13.826$, $p = .032$). Note that some participants indicated superiority for five monitors, a condition that was not included in this study[14], and some indicated superiority for more than one condition (three or four; four, five, or six).

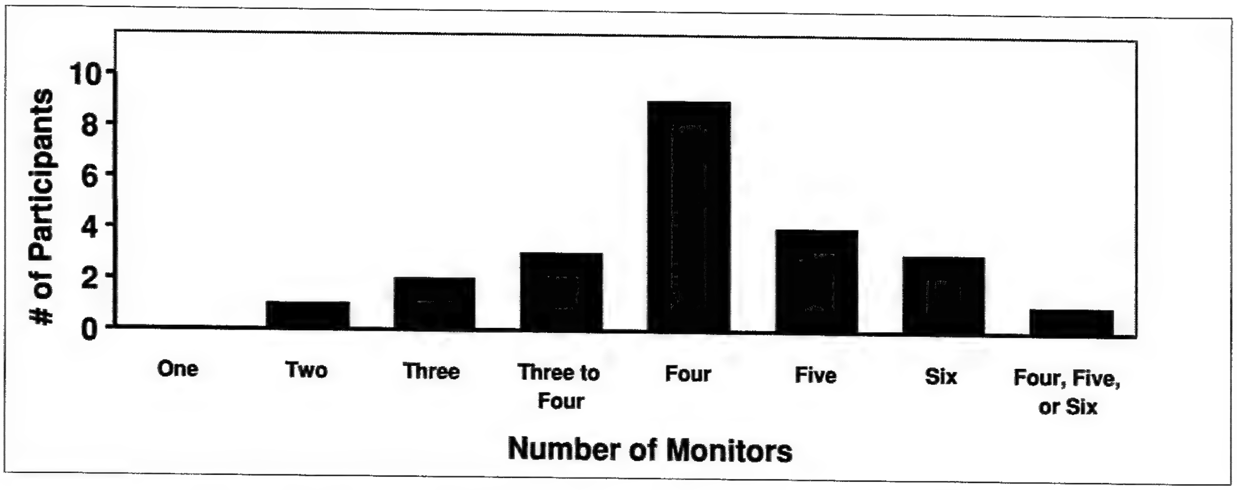

Figure 3. Preferred number of monitors indicated by participants in the Producer experiment.

Situation Awareness Questions—E-mail

Speed and accuracy of responses to e-mail inquiries for the five different monitor conditions are shown in Figure 4.

---

[14] One reason for this type of response may have been the wording of the question asked to participants. Specifically, they were asked "What is the number of monitors [not the monitor *condition*] that you believed best supported your tasks?" Because they could refigure the information in all conditions in the experiment (except the one-monitor condition), participants could have used fewer monitors than they had access to (e.g., they may have only used four or five monitors in the six-monitor condition).

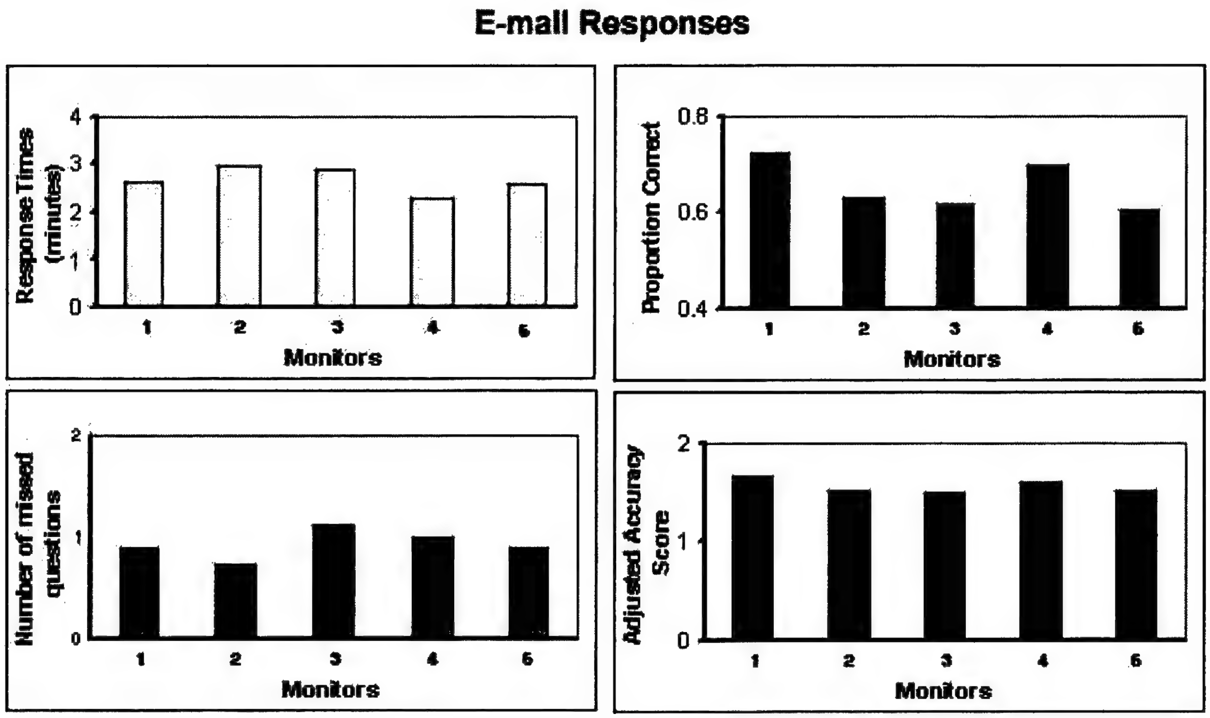

Figure 4. Speed, accuracy, misses, and adjusted scores for e-mail inquiry responses for the different monitor conditions in the Producer experiment.

Response Times. The slowest average response time was associated with the two-monitor condition (M = 2.96, SD = 2.12), followed closely by the three-monitor condition (M = 2.94; SD = 2.76). The fastest average response time was associated with the four-monitor condition (M = 2.29, SD = .96). However, there was no significant difference across conditions (F (2.71, 62.33) = .544, p = .636; partial η^2 = .023). Post hoc testing using the conservative Sidak test did not reveal any significant pairwise differences.

Proportion Correct. The highest mean proportion correct was associated with the one-monitor condition (M = .70, SD = .28), followed closely by the four-monitor condition (M = .72, SD = .21), and the lowest mean proportion correct was associated with the six-monitor condition (M = .60, SD = .25). There was no significant difference across conditions (F (4, 116) = 1.275, p = .284; partial η^2 = .042). The arcsine transformation was also not significant (F (4, 116) = 1.289, p = .278; partial η^2 = .043). Post hoc testing for the untransformed data using the Sidak test did not reveal any pairwise differences.

Misses. The highest mean for misses was associated with the three-monitor condition (M = 1.10, SD = 1.32), and the lowest mean for misses was associated with the two-monitor condition (M = .77, SD = .69). There was no significant difference across conditions (F (2.227, 99.773) = .647, p = .585; partial η^2 = .022). The degrees of freedom are adjusted given the violation of the sphericity assumption. Moreover, the Friedman test (χ^2 (4) = 791, p = .94) was not significant. Post hoc testing using the Sidak test did not reveal any significant pairwise differences.

Adjusted Accuracy Scores. The highest mean score was associated with the one-monitor condition (M =1.66, SD = .23), followed by the four-monitor condition (M = 1.59, SD = .29), and the lowest mean was obtained for the three-monitor condition (M = 1.50, SD = .32). There was no significant

difference across conditions (F (4, 116) = 1.75, p = .144; partial η^2 = .057). Post hoc testing using the Sidak test did not reveal any significant pairwise differences.

Situation Awareness Questions—Chat

Speed and accuracy of chat inquiry responses for the five different monitor conditions are shown in Figure 5.

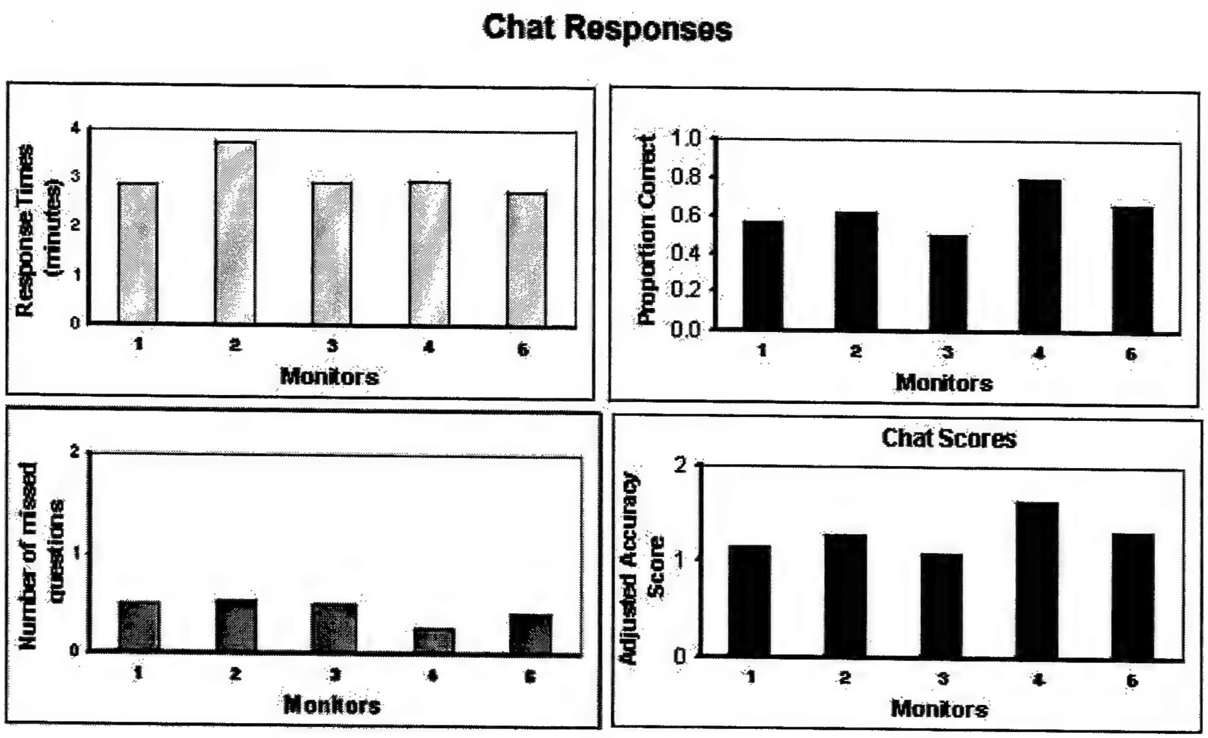

Figure 5. Speed, accuracy, misses, and adjusted scores for chat inquiry responses for the different monitor conditions in the Producer experiment.

Response times. The slowest mean response time was associated with the two-monitor condition (M = 3.75, SD = 2.86), and the fastest mean response time was associated with six monitors (M = 2.76, SD = 1.63). There were no chat questions presented in one block of the experiment (yielding missing data for every participant in one condition); therefore, repeated measures analysis of variance could not be conducted. Instead, paired t-tests were performed (e.g., one-monitor vs. two-monitor condition, two-monitor vs. six-monitor condition, etc.); however, none of those exploratory tests obtained significance.

Proportion Correct. The highest mean proportion correct was associated with the four-monitor condition (M =.79, SD = .36), and the lowest mean proportion correct was associated with the three-monitor condition (M = .50, SD = .47). For the same reasons as above, paired t-tests were conducted. There was a significant difference between the one-monitor and four-monitor conditions (t (17) = -3.42, p = .003 (95% confidence interval of the difference = [-.584, -.1386] for both the untransformed and arcsine transform data)).

Misses. The highest mean for misses was associated with the two-monitor condition (M = .54, SD = .78), followed closely by the one-monitor and three-monitor conditions (M =.50, SDs = .59 and .51, respectively). The lowest mean was associated with the four-monitor condition (M = .25, SD = .44).

As with the prior analyses, the pattern of missing data was severe enough to preclude simultaneous estimation of all five conditions. Thus, a Wilcoxon Signed Ranks Test was used to test the significance of paired conditions. None of these comparisons was significant.

Adjusted Accuracy Scores. The highest mean score was associated with the four-monitor condition (M =1.63, SD = .65), and the lowest mean was obtained for the three-monitor condition (M = 1.08, SD = .88). A Wilcoxon Signed Ranks Test was used to test the significance of certain paired conditions. Significance was found for the three-monitor vs. four-monitor condition (z = -2.02, p = .044) and for the one-monitor vs. four-monitor condition (z = -2.52, p = .012).

Information Products—Product Accuracy

Inter-rater reliability. The intraclass correlation coefficient (ICC) across k = 3 raters for each scenario was calculated to assess inter-rater reliability.[15] The single measure ICC is the index reported in this study because it is generally more conservative than the average measure ICC. An ICC between .7 and .8 was considered acceptable, because it corresponds to the generally acceptable limits of internal consistency estimates.

Table 2 shows the range of the three Pearson correlation coefficients calculated for each scenario, the single measure ICC, and 95% confidence interval. In summary, the scenarios with the highest levels of inter-rater reliability per the single measure ICC were India and Korea. All others, except Bangladesh, were within acceptable parameters. Note that India not only had the highest ICC but also the tightest confidence interval.

Table 2. Inter-rater reliability for the product accuracy ratings.

Scenario	Pearson Correlation Coefficient Range	Single Measure ICC	95% Confidence Interval
Cambodia–Intel	.68 – .95	.8048	.6545 – .9042
Korea–Metoc	.76 – .87	.8153	.6847 – .9037
Bangladesh–Logistics	.70 – .82	.6805	.4990 – .8204
China–Air Defense	.81 – .87	.7182	.5465 – .8457
India–Force Protect	.69 – .76	.8843	.7833 – .9427

Scores. Figure 6 shows the composite product accuracy scores for the different monitor conditions (average across three raters). The highest mean score was associated with the six-monitor condition (M = 11.78, SD = 4.31), and the lowest with the three-monitor condition (M = 8.56, SD = 4.43). There was a significant difference across conditions (F (4, 64) = 2.563, p = .047; partial η^2 = .138). However, post hoc testing using the Sidak test did not reveal any pairwise differences.

[15] The advantages of the intraclass correlation are such that it teases out variance due to the judges, which the Pearson correlation coefficient fails to do (Shrout & Fleiss, 1979). For this analysis, the two-way random effects model (i.e., Model 2) with an emphasis on rater consistency (as opposed to absolute agreement) was used.

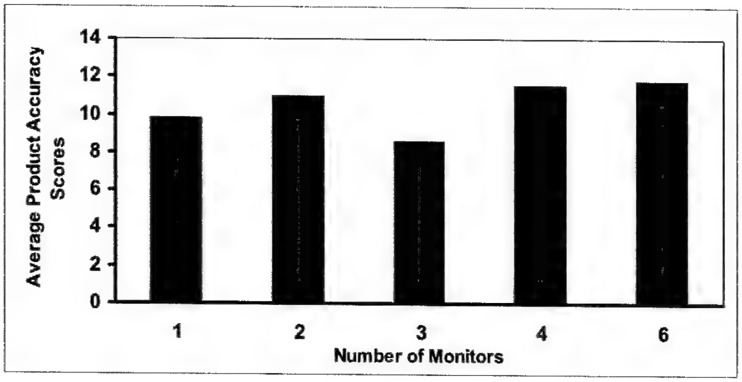

Figure 6. The average composite product accuracy scores for the different monitor conditions in the Producer experiment.

Information Products—Whole Product Quality

Inter-rater reliability. In examining inter-rater reliability for the whole product quality scores, the use of the ICC was inappropriate given the extent of missing data and, most prominently, the use of a three-point ordinal scale. The Spearman rank-order correlation coefficient (r_s) was used instead to calculate the association across all $k = 3$ raters for each scenario (i.e., Cambodia, Korea, Bangladesh, China, and India). Table 3 shows the Spearman rank order correlations obtained for each summary. In contrast with the product accuracy scores, inter-rater reliability for the quality scores was generally not very high.

Table 3. Inter-rater reliability for the whole product quality ratings.

Scenario	Right Scores			Wrong Scores			Processing Scores		
	TR1/ TR2	TR1/ SME	TR2/ SME	TR1/ TR2	TR1/ SME	TR2/ SME	TR1/ TR2	TR1/ SME	TR2/ SME
Cambodia–Intel	.647*	.116	-	.156	.600	-	.791*	.053	-
Korea–Metoc	.079	.280	-.465	.309	.396	.249	.569*	.060	-.215
Bangladesh–Logistics	.473	.688*	-	.003	.200	-	.389	.320	-
China–Air Defense	.491	-.018	.051	.582*	.675*	.866*	.487*	.252	.683*
India–Force Protect	.161	-.244	.025	.741*	.890*	.721*	-.070	.236	.167

* Statistically significant (alpha = .05); a blank cell indicates that two raters did not rate the same products within a scenario.

TR = trained rater; SME = subject matter expert. Note that the TRs and SMEs were not the same across the scenarios.

Scores. Figure 7 shows the average whole product accuracy scores for each of the three scales, across the different monitor conditions. For "right"scores, the highest mean was associated with the four-monitor condition ($M = 1.84$; $SD = .40$) and the lowest with the two-monitor condition ($M = 1.7$; $SD = .43$). For "wrong" scores, the highest mean was associated with the six-monitor condition ($M = 2.42$; $SD = .62$) and the lowest with both the one-monitor condition ($M = 2.25$; $SD = .62$) and the four-monitor condition ($M = 2.25$; $SD = .60$). For "processing" scores, the highest mean was associated with the two-monitor condition ($M = 1.53$, $SD = .59$) and the lowest with the four-monitor

condition ($M = 1.31$, $SD = .40$). No significant differences were found across conditions for any of the scales (for right scores, $F(4, 76) = .365, p = .833$, partial $\eta^2 = .019$ (Friedman test: $\chi^2(4) = 3.17$, $p = .53$) for wrong scores; $F(4, 76) = .415, p = .798$, partial $\eta^2 = .021$ (Friedman test $\chi^2(4) = 1.82, p = .769$) for processing scores; and $F(4, 76) = .977, p = .425$, partial $\eta^2 = .049$ (Friedman test $\chi^2(4) = 3.77, p = .438$))[16]

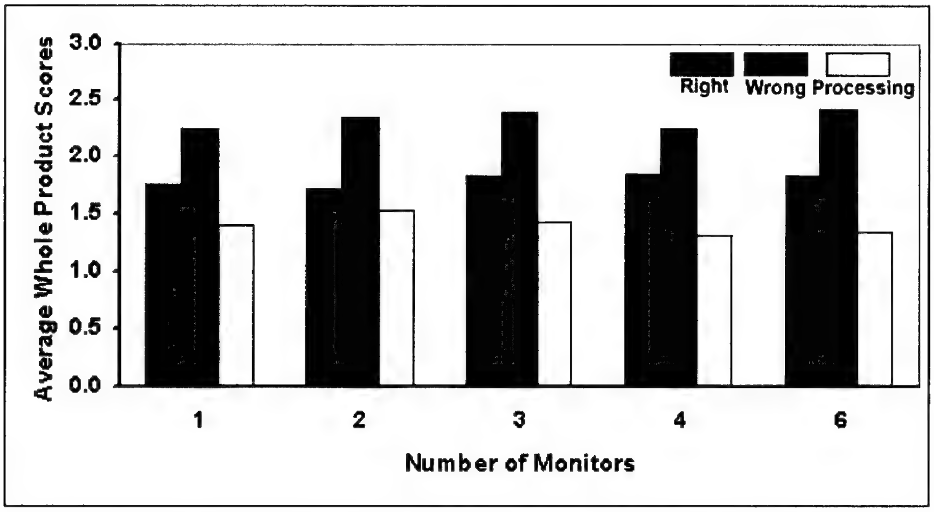

Figure 7. The average whole product quality scores for the different monitor conditions in the Producer experiment.

DISCUSSION

In order to determine the optimum number of monitors to support warfighters in the producer role, we must consider performance across *all* dependent measures—both separately and in aggregate (the "big picture"). Table 4 shows the monitor conditions that produced the best performance (highest preference, fastest response times, highest accuracy) and worst performance (lowest preference, slowest response times, lowest accuracy) for each of the dependent variables in the Producer experiment.[17]

---

[16] There did appear to be sufficient variation, across conditions, to use repeated measures analysis. However, given the inherent ordinal metrics of the data, the Friedman two-way analysis of variance by ranks, a test of medians for matched samples, was the nonparametric technique of choice.

[17] Those variables with statistically significant differences across monitor conditions (i.e., alpha = .05) are indicated as such. Close second best (or close second worst) condition are reported in parentheses.

Table 4. The monitor conditions that produced the best and worst performance for each of the dependent variables in the Producer experiment.

Producer Experiment			
Task	**Measure**	**"Optimum" Number of Monitors (2^{nd} best, if close)**	**"Worst" Number of Monitors (2^{nd} worst, if close)**
All	Preference	Four *	One (Two)
E-mail	Reaction Time	Four	Two (Three)
	Proportion Correct	One (Four)	Six
	Misses	Two	Three
	Adjusted Accuracy Scores	One (Four)	Three (Two)
Chat	Reaction Time	Six	Two
	Proportion Correct	Four**	Three
	Misses	Four	Two (One, Three)
	Adjusted Accuracy Scores	Four***	Three
Information Product	Quality	Six	One
	Accuracy	Three	One

* significant differences across conditions, alpha = .05

** significantly better than One Monitor, alpha = .05

*** significantly better than Three Monitors and Four Monitors, alpha = .05

Performance in the different monitor conditions was rank ordered for each of the dependent variables as follows: 1st = best performance (e.g., fastest response times, highest scores), and 5th = worst performance). To assess inter-test reliability, the nonparametric Kendall coefficient of concordance W was used to determine the extent of association/agreement among rankings. A chi-square distribution was used to test significance for Kendall's W. A significant value of W indicates that there is symmetry (i.e., agreement) across the rankings. The omnibus analysis yielded significant concordance for all conditions ($\chi2$ (4) = 12.5, p = .014) with the four-monitor condition yielding the highest rank and the three-monitor condition yielding the lowest rank. A significant difference was found when comparing the four-monitor condition to the three-monitor condition ($\chi2$ (1) = 7.36, p = .007). No significant differences were found when comparing the four-monitor condition to the one monitor condition ($\chi2$ (1) = .818, p = .366), the two-monitor condition ($\chi2$ (1) = 4.46, p = .035), or the six-monitor conditions ($\chi2$ (1) = .82, p = .366).

In summary, the pattern of results points to four monitors as the optimum number for the tasks performed in the Producer experiment. As reported in Table 4, overall, the four-monitor condition supported the best performance for e-mail and chat tasks and was the condition participants thought best supported their task. The poorest overall performance was yielded by the two-monitor and three-monitor conditions. However, most of the analyses used when comparing performance across the conditions did not yield statistically significant results. This is most likely due to limitations placed on the experimental design due to the nature of an LOE (such as low power, limited task time) and suggests the need for further research before any strong conclusions can be made.

CONSUMER EXPERIMENT

In the Consumer experiment, the participant assumed the role of a CJTF/senior commander/Battle Watch Captain (BWC) monitoring a fictional operational situation. Their primary task was to acquire and maintain situation awareness by monitoring available information sources and a tactical display. In the Consumer experiment, various kinds of information were presented via a website. While using these resources, participants were also required to monitor incoming information and answer questions using chat and e-mail. In each block of the experiment, they monitored a new scenario taking place in one of five geographic locations.[18] Participants performed their tasks using one, two, three, four, or six monitors. (See [5])

As in the Producer experiment, the number of e-mail or chat inquiries requiring a response, as well as the number of e-mails and chat rooms to monitor, was based on the results of a survey sent to fleet users prior to the experiment. Performance in the experiment was based on the speed and accuracy of participant responses to the chat and e-mail questions (which were used to assess participants' situation awareness) and the number of accesses to web pages they made.

METHOD

Participants

Thirty participants each served in a 2½ hour session. Between one and five participants were run in each session concurrently. Participants were instructors, students, or support personnel at the Naval War College. Demographic and computer experience information, collected at the conclusion of the experiment, is shown in Table 1. All participants were active or retired military (participants that reported rank as "civilian" were retired military), primarily Navy, with an average of 19.9 years of service.

Design

The design of the experiment was within-subject. Every participant served in each of the five display conditions (1, 2, 3, 4, and 6 monitors) presented in five 20-minute blocks. A Latin square was used to counterbalance the order of display conditions across participants. The five scenarios (Luzon, Aceh, Java, Mindanao, and Visayas) were presented in the same order for every participant.

Procedure

Each participant performed all tasks using a K-Desk with 1, 2, 3, 4, or 6 windows activated. The software applications (e.g., *MS Chat, MS Outlook, MS Internet Explorer,* and *Geoplot*[19]) were displayed and configured on the desktop by the experimenter at the beginning of each block (see Figure 8). Participants could move applications and documents into any active window of the K-desk as desired.

---

[18] Scenarios took place in Luzon, Aceh, Java, Mindanao, and the Visayas.

[19] Chat®, Outlook®, and Internet Explorer® are registered trademarks of the Microsoft corporation. Geoplot is a software product of Pacific Science and Engineering Group, Inc.

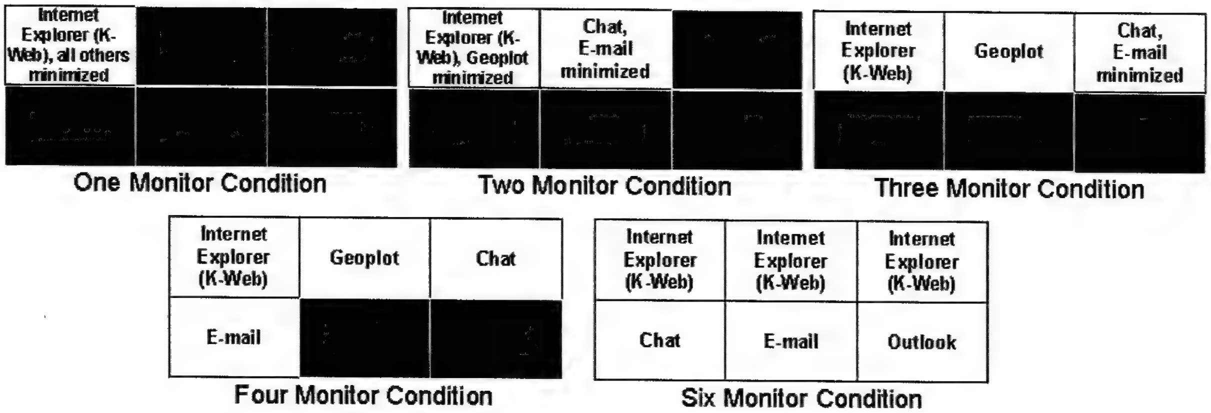

Figure 8. Initial display configurations for the monitor conditions in the Consumer experiment.

At the beginning of the experimental session, the participant received approximately 20 minutes of training. This training included an overview of the experimental task, followed by hands-on instruction on the following: (1) browsing the web, (2) configuring the desktop, (3) using the tactical display (*Geoplot)*, and (4) switching between software applications. Participants were given a chance to ask questions during and after the training session.

At the beginning of each block, a description of the scenario was read to the participants. They were then told that they had 20 minutes to monitor the situation. During those 20 minutes, participants performed four tasks concurrently:

1. *Browsing the Web.* Participants were given access to a Knowledge Web (K-Web) website (see Rogers, et al., 2002 for more detailed information), which they could browse to acquire situation awareness about the scenario. The K-Web consisted of five "summary" web pages, each authored by a different FCC. Summary pages included color-coded status information related to the scenario and hyperlinks to more detailed products. The summary pages could be accessed from an "overview" page, which provided integrated status information from all summary pages. Example overview and summary pages are shown in Figure 9.

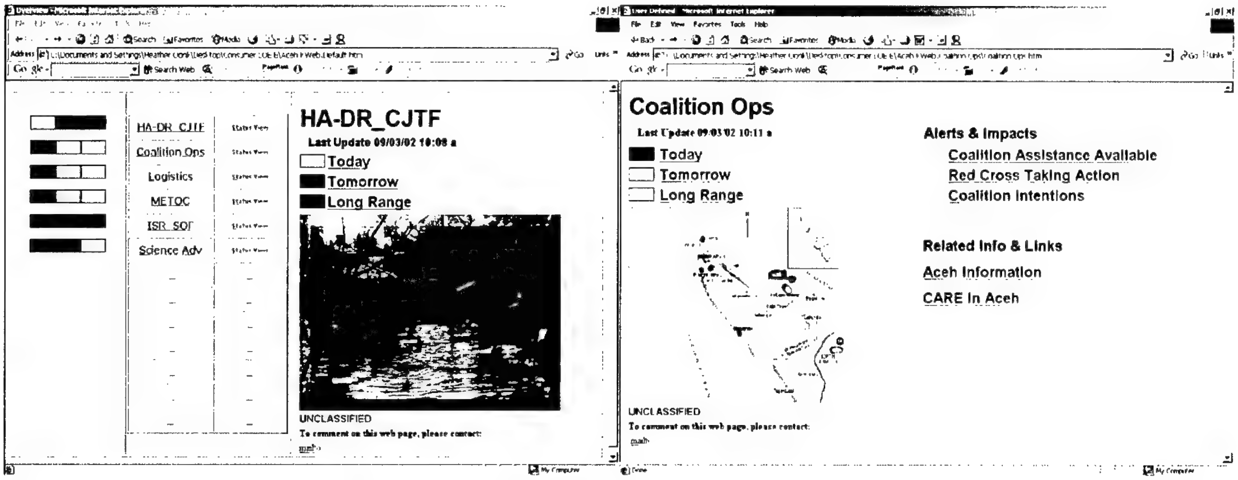

Figure 9. Example K-Web overview page (left) and summary page (right).

2. *Monitoring a tactical display.* Participants were provided a basic tactical picture by the *Geoplot* software (see Figure 10). This application displayed a map populated with air and sea surface tracks represented by standard Navy Tactical Data System (NTDS) symbols. Clicking on the tracks allowed users to get amplifying information (track number, bearing, range, course, speed, name, assets, and assigned unit).

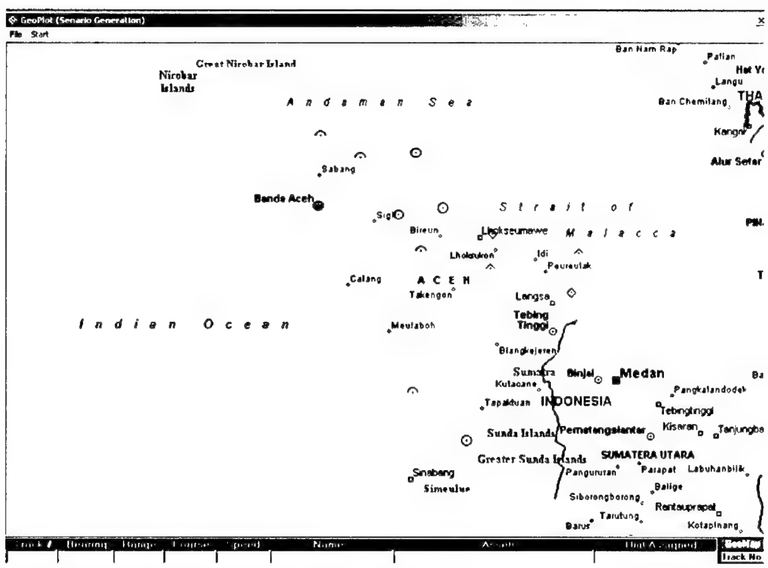

Figure 10. Geoplot tactical display.

3. *Monitoring and responding to chat messages.* Participants were required to monitor five chat rooms and respond to any questions presented in them. During the course of the scenario, they received seven chat messages. Four or five of the messages contained new information about the scenario; the remainder contained a question posed by Higher Authority. They were told that their performance would be rated on the speed and accuracy of their responses to these questions.

4. *Monitoring and responding to e-mail messages.* Participants were required to read and respond to seven e-mails. One or two of the e-mails contained new information about the scenario; the remainder contained a question posed by Higher Authority. They were told that their performance would be rated on the speed and accuracy of their responses to these questions.

In total, participants received six pieces of new information and eight questions over the course of the scenario. The presentation of these was split between the chats and e-mails sent to the participants by confederates—the timing and text of these messages was determined by a script. Each experiment participant was assigned a confederate who played the role of Higher Authority and various other members of the command/battlegroup. The confederates were only allowed to send messages according to the script—they did not respond to communications initiated by the participant. Answers to the questions were available in the K-Web, the tactical display, or in earlier chats or e-mails.

At the end of the experiment, participants filled out a demographic information form and were asked to indicate which monitor condition they thought best supported their tasks.

DATA ANALYSIS

The same dependent variables that were analyzed for the Producer experiment were analyzed for the Consumer experiment, with the following exceptions:

- *Web Access.* The number of web accesses (total number of links clicked) was an added variable.
- *Information product quality and accuracy* were *not* analyzed (because no information products were created).

RESULTS

Appendix C contains tables of the automated data presented in the figures in the Results sections.

Preference

As shown in Figure 11, most participants ($n = 10$) indicated that they thought the four-monitor condition best supported their tasks. Chi square analysis, however, revealed no significant difference between the number of people who indicated a preference across the monitor conditions ($\chi^2(6) = 4.621, p = .328$).

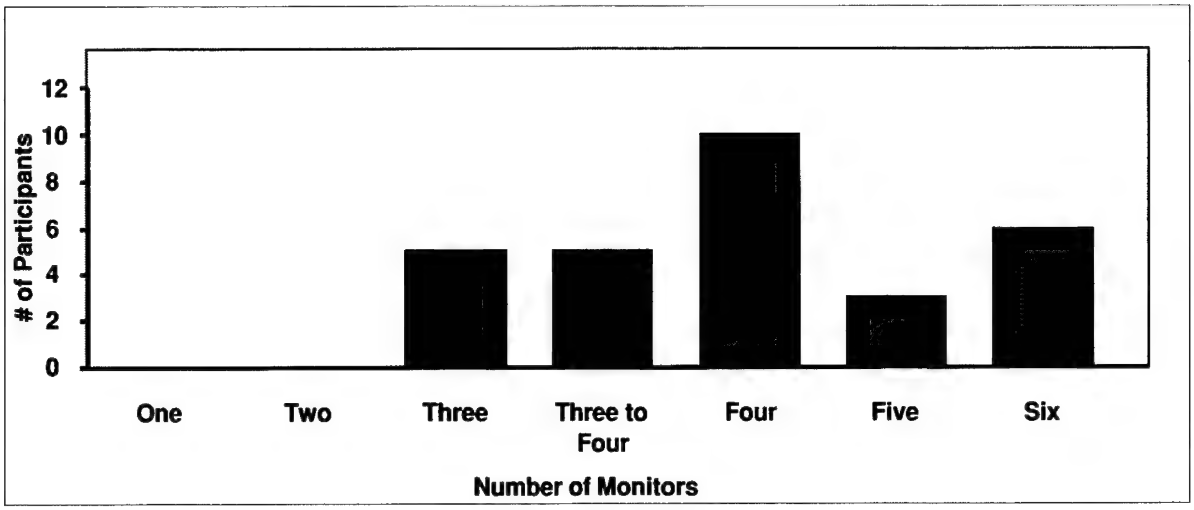

Figure 11. Preferred number of monitors indicated by participants in the Consumer experiment.

Situation Awareness Questions—E-mail

Speed and accuracy of responses to e-mail inquiries for the five different monitor conditions are shown in Figure 12.

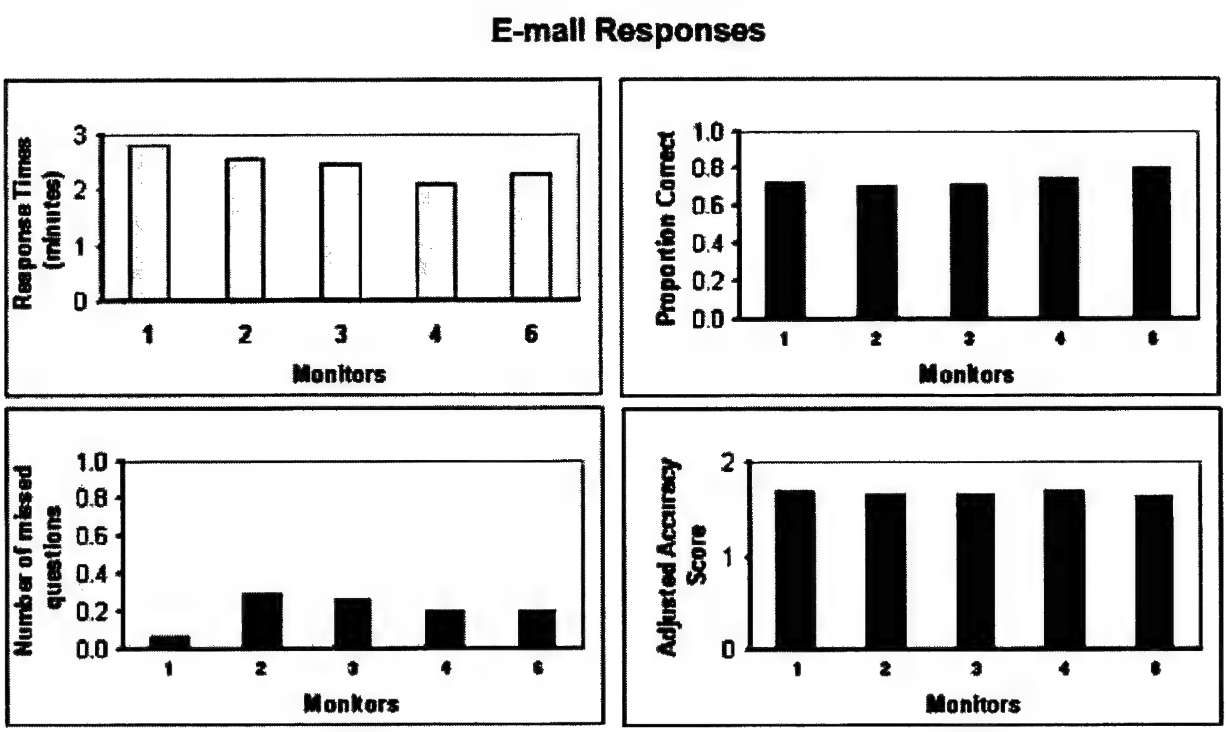

Figure 12. Speed, accuracy, misses, and adjusted scores for e-mail inquiry responses for the different monitor conditions in the Consumer experiment.

Response Times. The slowest mean response time was associated with the one-monitor condition (M = 2.83, SD = 1.55) and the fastest mean response time with the four-monitor condition (M = 2.13,

$SD = 1.04$). There was no significant difference across conditions ($F(4, 116) = 1.779, p = .138$; partial $\eta^2 = .058$). Post hoc testing using the conservative Sidak test did not reveal any pairwise differences.

Proportion Correct. The highest mean proportion correct was associated with the six-monitor condition ($M = .80, SD = .19$) and the lowest mean proportion correct with two-monitor condition ($M = .70; SD = .23$). There was no significant difference across conditions ($F(4, 116) = 1.058, p = .381$; partial $\eta^2 = .035$). The arcsine transformation was also not significant ($F(4, 116) = .844, p = .500$; partial $\eta^2 = .028$). Post hoc testing for the untransformed data using the Sidak test did not reveal any pairwise differences.

Misses. The highest mean for misses was associated with the two-monitor condition ($M = .30, SD = .54$), and the lowest mean was associated with the one-monitor condition ($M = .067, SD = .25$). There was no significant difference across conditions ($F(4, 116) = .921, p = .455$; partial $\eta^2 = .031$). Moreover, the Friedman test ($\chi^2(4) = 4.10, p = .393$) was not significant. Post hoc testing using the Sidak test did not reveal any pairwise differences.

Adjusted Accuracy Scores. The highest mean score was associated with the four-monitor condition ($M = 1.681; SD = .21$ and one-monitor condition ($M = 1.680; SD = .23$). The lowest mean was obtained for the six-monitor condition ($M = 1.63, SD = .22$). There was no significant difference across conditions ($F(4, 116) = .4133, p = .799$; partial $\eta^2 = .014$). Post hoc testing using the Sidak did not reveal any pairwise differences.

Situation Awareness Questions—Chat

The response times and accuracy to answer the situation awareness questions presented and answered in chat for the five different monitor conditions are shown in Figure 13.

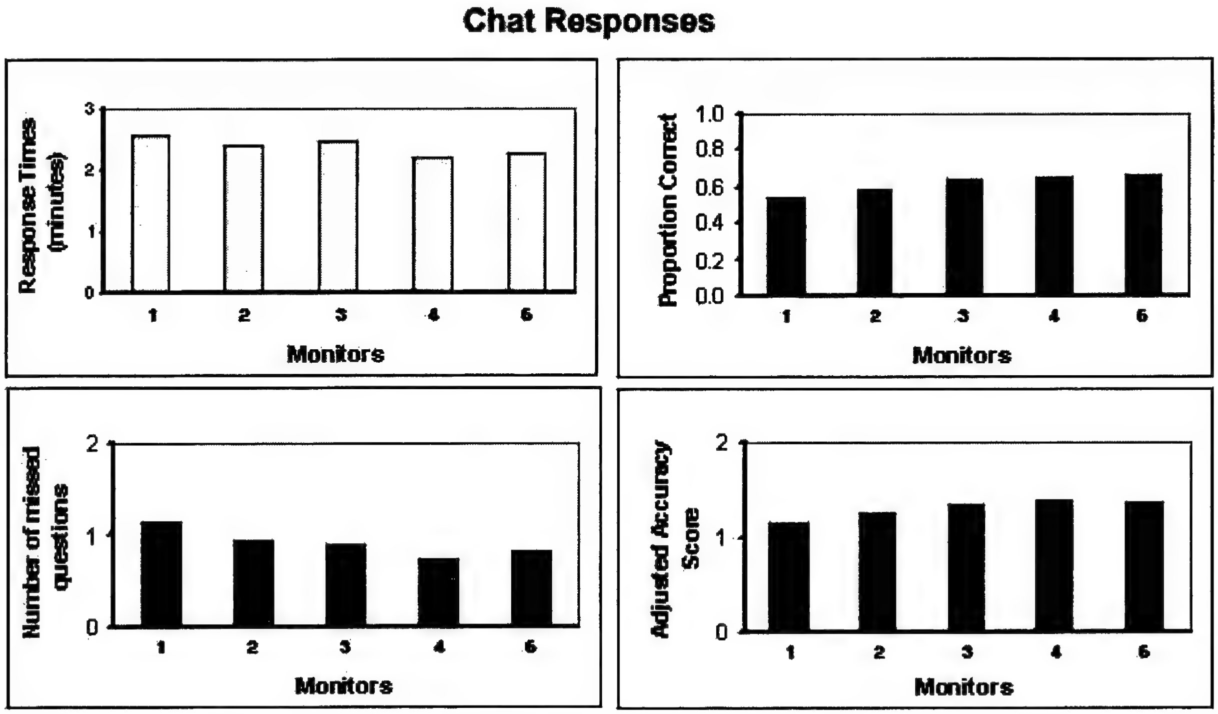

Figure 13. Speed, accuracy, misses, and adjusted scores for chat inquiry responses for the different monitor conditions in the Consumer experiment.

Response Times. The slowest mean response time was associated with the one-monitor condition (M = 2.59, SD = 1.79), and the fastest was associated with the four-monitor condition (M = 2.19, SD = 1.33). There was no significant difference across conditions (F (4, 92) = .277, p = .892; partial η^2 = .012). There were no significant pairwise differences per post hoc tests.

Proportion Correct. The highest mean proportion correct was associated with the six-monitor condition (M = .66; SD = .28), followed closely by the four-monitor condition (M = .66, SD = .32). The lowest mean was associated with the one-monitor condition (M = .53, SD = .29). There was no significant difference across conditions (F (4, 116) = 1.274, p = .284; partial η^2 = .042). The arcsine transformation was also not significant (F (4, 116) = 1.512, p = .203; partial η^2 = .05). Post hoc testing for the untransformed data and transformed data, using the Sidak test did not reveal any pairwise differences.

Misses. The highest mean for misses was associated with the one-monitor condition (M = 1.13, SD = .94), and the lowest mean was associated with the four-monitor condition (M = .73, SD = .91). There was no significant difference across conditions (F (4, 116) = .892, p = .471; partial η^2 = .03). Moreover, the Friedman test (χ^2 (4) = 3.53, p = .474) was not significant. Post hoc testing using the Sidak test did not reveal any pairwise differences.

Adjusted Accuracy Scores. The highest mean score was associated with the four-monitor condition (M = 1.39; SD = .62), with the six-monitor condition closely following (M = 1.38; SD = .53). The lowest mean was obtained for the one-monitor condition (M = 1.14; SD = .59). There was no significant difference across conditions (F (4, 116) = 1.136, p = .343; partial η^2 = .038). Post hoc testing using the Sidak did not reveal any significant pairwise differences.

Web Accesses

The mean number of web accesses (mouse clicks on hyperlinks to access a web page) during a block for the five different monitor conditions are shown in Figure 14.

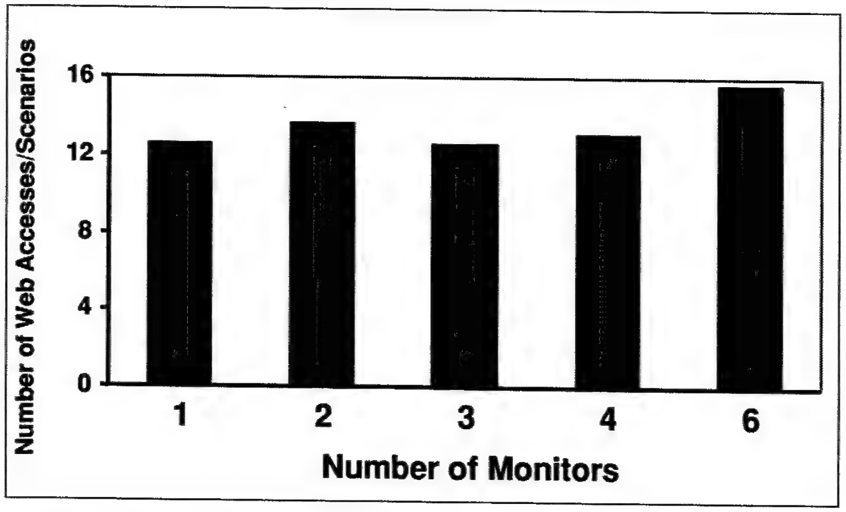

Figure 14. Mean number of web accesses during a block in the Consumer experiment.

The highest mean number of web accesses was associated with the six-monitor condition (M = 13.94, SD = 4.28), and the lowest mean was associated with the one-monitor (M = 12.53, SD = 5.26) and three-monitor conditions (M = 12.53, SD = 7.32). There was no significant difference across conditions (F (4, 116) = 2.38, p = .056; partial η^2 = .076). Post hoc testing using the conservative Sidak test revealed the following significant pairwise difference: one-monitor vs. six-monitor conditions (p = .034, 95% CI = [.152, 5.982]).[20]

DISCUSSION

As in the Producer experiment, we considered performance across *all* dependent measures—both separately and in aggregate. Participants indicated that they thought the four-monitor condition best supported their task. Table 5 shows the monitor conditions that produced the best performance and worst performance for each of the dependent variables in the Consumer experiment.

Each of the conditions was compared to the four-monitor condition (using Kendall's W, alpha = .0125). There was significant agreement on the rankings when comparing the four-monitor condition with the one-monitor condition (χ^2 (1) = 6.4, p = .011), the two-monitor condition (χ^2 (1) = 6.4, p = .011), and the three-monitor conditions (χ^2 (1) = 10.0, p = .002), with the four-monitor condition yielding a higher ranking. However, there was no significant agreement on the rankings when comparing the six- and four-monitor conditions (χ^2 (1) = 1.0, p = .317).

Each of the conditions was compared to the four-monitor condition (using Kendall's W, alpha = .0125). There was significant agreement on the rankings when comparing the four-monitor condition with the one-monitor condition (χ^2 (1) = 6.4, p = .011), the two-monitor condition (χ^2 (1) = 6.4, p = .011), and the three-monitor conditions (χ^2 (1) = 10.0, p = .002), with the four-monitor condition

[20] This result should be interpreted with caution, however, as the omnibus test was not statistically significant.

Table 5. The monitor conditions that produced the best and worst performance for each of the dependent variables in the Consumer experiment.

Consumer Experiment			
Task	**Measure**	**"Optimum" Number of Monitors (2nd best, if close)**	**"Worst" Number of Monitors (2nd worst, if close)**
All	Preference	Four	One / Two
E-mail	Reaction Time	Four	One
	Proportion Correct	Six	Two
	Misses	One	Two
	Adjusted Accuracy Scores	Four	Six
Chat	Reaction Time	Four	One
	Proportion Correct	Six (Four)	One
	Misses	Four	One
	Adjusted Accuracy Scores	Four	One
Monitoring	Web Accesses	Six*	One / Three
*significantly better than One Monitor, alpha = .05			

yielding a higher ranking. However, there was no significant agreement on the rankings when comparing the six- and four-monitor conditions (χ^2 (1) = 1.0, p = .317).

As in the Producer experiment, the pattern of results points to four monitors as the optimum number for the tasks performed in the Consumer experiment. As reported in Table 5, e-mail responses and chat were best supported by the four-monitor condition. The number of web accesses was highest in the six-monitor condition. Although this is only one measure related to a monitoring task (our data collection procedures did not allow us to collect eye and mouse movement data[21]), it suggests that the important task of monitoring the operational situation was best supported in this condition. Overall, the one-monitor condition yielded the poorest performance across all tasks. As in the Producer experiment, these results suggest the need for further research before any strong conclusions can be made.

[21] For example, users of more monitors have the advantage of the ability to visually scan more screen real estate simultaneously, suggesting improved performance in monitoring tasks. However, large eye and head movements may produce a deficit in performance (see Introduction for a review of these issues).

GENERAL DISCUSSION

SUMMARY

Overall, the four-monitor condition supported the best performance in both the Producer and Consumer experiments. When comparing the different dependent measures in each experiment, a similar pattern emerged, i.e., there was a fairly consistent trend of "best" to "worst" monitor conditions. Figure 15 shows the rankings, averaged across all the dependent variables measured in the two experiments. (Note that a lower value ranking indicates better performance.) According to this figure, the results of the Consumer experiment support the hypotheses listed in the introduction. Performance improved as the number of monitors increased and performance tended to improve only up to a point of diminishing returns. In the Producer task, however, there was no increase in performance with two or three monitors when compared to four. Further, *overall* performance did not asymptote for fewer monitors in the Producer experiment than the Consumer experiment as we had predicted it would. An interesting finding from both experiments was that the optimum condition in terms of user preference was also the one in which participants performed the best—this is not always the case in applied experiments (Andre & Wickens, 1995; Bailey, 1993).

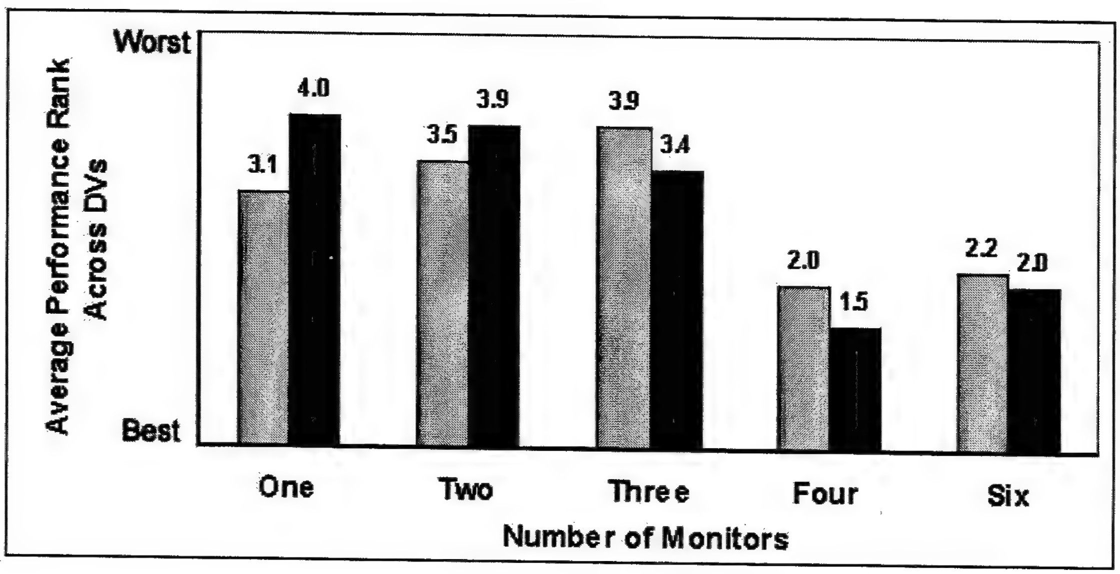

Figure 15. Average ranks derived when comparing performance across all dependent variables measured in the Producer and Consumer experiments.

Conclusions based on the pattern of results shown in Figure 15 can only be made if it is believed that each of the dependent variables should be weighed equally. The results of the experiments suggest that the optimum number of monitors is task dependent—different for the overall producer and consumer tasks, but also for the subtasks that they comprise. For example, in the Consumer task, if emphasis is placed on monitoring the operational mission, six monitors supported superior performance. This is an important distinction that should be taken into account in the design of workstations.

Although there are patterns in the data from the above experiments that point to an "optimum" number of monitors for the tasks examined, they also suggest the need for further research. The nature of an LOE, which by definition is a "small scale" study, restricted the number of participants

that could serve in the experiment (and thus statistical power) and the length of time for each experimental session.[22] More refined comparisons should be made before the results of the above experiments are applied to any workstation design. Future studies should compare performance in producer and consumer tasks across fewer conditions, using more realistic and sensitive tasks by giving the participants more information to monitor/integrate and longer blocks of time to perform their tasks. Producer studies should focus on the comparison between three, four and perhaps more monitors while consumer studies should focus on performance with four to six and even more monitors.

FURTHER RESEARCH

As discussed in the introduction, these results should be treated as very situation-dependent and only applied to the specific warfighter context examined. Further screen real estate is not a stand-alone concern with respect to performance and other issues, such as configuration, are likely to interact with number of monitors. The experiments reported in this paper provide a "first step" in the examination of issues related to multiple monitors—helping to define the parameters under which other, more interesting, multiple configuration-related variables can be pursued.

The configuration of information and applications used by the warfighter is an issue that should be examined in future studies. These studies should examine not only the configuration of monitors, but more importantly, the configuration of information within the workstation display. Many questions can be examined in such studies, including:

- What types of tasks are best supported by multi-monitor displays?
- How does the number and configuration of information have an effect on cognitive workload?
- Which display configuration best supports the cognitive processes involved in warfighter tasks, such as monitoring, decision-making, data integration, pattern recognition, and attention management?
- What are the costs and benefits associated with different degrees of user control over display configuration?

As with the experiments reported here, however, such studies should be designed with realistic tasks in mind, and their results should be applied with context-dependency taken into account.

Further, it could prove highly desirable to incorporate gaze (eye and head) monitoring instrumentation into this research as a means of increasing the detail of analysis of how the displays are being used. Clearly there are going to be trade-offs in the location and organization of information that would be much better understood if we could know where decision-makers were looking as they performed different types of tasks (see, for example, Morrison, et al., 1997). Data collected using such monitoring instrumentation would be valuable if used to augment the type of data collected in the current experiment. For example, it could provide us additional data related to Web browsing (see Footnote 21).

RECOMMENDATIONS

Warfighters must have enough monitors to support their task, but not so many that they are overloaded by information. Overestimation of monitors needed by fleet users can result in performance decrements and unnecessary fiscal costs. Underestimation, on the other hand, may also result in undesirable performance costs in terms of speed and quality of decision-making. Ideally,

[22] Data collection took place over a single 4-day period (16–19 September 2002).

research in this direction should allow us to make value estimates per monitor, i.e., best return on investment with respect to performance. Based on the findings of the two experiments reported in this paper:

- Four monitors are recommended for producer tasks similar to the ones examined here, i.e., involving creation of information products through the integration of multiple sources of information, concurrent with monitoring of incoming information and responding to requests for information.
- At least four, and up to six or more monitors are recommended for consumer tasks similar to the ones examined here, i.e., monitoring of an operational situation concurrent with monitoring of incoming information and responding to requests for information.
- Research is needed to compare performance in these tasks in a more robust manner, preferably including instrumentation to allow a more detailed analysis in how to optimally configure and present information within the multi-monitor workstations.
- Research is necessary to examine other issues related to multi-monitor displays used by the warfighter.

REFERENCES

Andre, A. D. and Wickens, C. D. 1995. "When users want what's NOT best for them." *Ergonomics in Design*, October, 10-14.

Baddeley, A. D. 1986. *Working memory*. Oxford: Oxford University Press.

Bailey, R. W. 1993. "Performance vs. preference." *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society. pp. 282-286.

Card, S. K. and Henderson, A. 1987. "A multiple virtual workspace interface to support user task." In *Proceedings of ACM CHI+GI '87 Conference on Human Factors in Computing Systems and Graphics Interface*. New York, NY, ACM Press, pp.53-59.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. 2003. *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd Ed.). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. R. 1995. "Toward a theory of situational awareness in dynamic systems." *Human Factors*, 37, 32-64.

Fitts, P. M. 1954. "The information capacity of the human motor system in controlling the amplitude of movement." *Journal of Experimental Psychology,* 47, pp. 381-391.

Gillan, D. J., Holden, K., Adam, S. Rudisill, M., and Magee, L. 1990. "How does Fitts' Law fit pointing and dragging." *Proceedings of ACM CHI'90 Conference on Human Factors in Computing Systems,* pp. 227-234.

Grudin, J. 2001. "Partitioning digital worlds: focal and peripheral awareness in multiple monitor use." *CHI 2001 Proceedings*, pp. 458-465.

Morrison, J. G., Marshall, S. P., Kelly, R. T., and Moore, R. A. 1997. "Eye tracking in tactical decision making environments: implications for decision support evaluation." In *Proceedings of the Third International Symposium on Command and Control Research and Technology*. National Defense University, June 17-20, 1997.

Oel, P., Schmidt, P., and Schmitt, A. 2001. "Time prediction of mouse-based cursor movements." *Proceedings of Joint AFIHM-BCS Conference on Human Computer Interaction IHM-HCI'2001*. Lille, France, pp. 37-40.

Oonk, H. M., Smallman, H. S., Moore, R. A., and Morrison, J. G. 2000. "Usage, utility, and usability of the Knowledge Wall during the Global 2000 War Game." SPAWAR Systems Center San Diego, CA: Technical Report 1861.

Oonk, H. M., Rogers, J. H., Moore, R. A., and Morrison, J. M. 2002. "Knowledge Web concept and tools: Use, utility and usability during the Global 2001 War Game." SPAWAR Systems Center San Diego, CA: Technical Report 1882.

Robinson, G. H. 1979. "Dynamics of the eye and head during movement between displays: A qualitative and quantitative guide for designers." *Human Factors,* 21, pp. 343-352.

Rogers, J. H., Oonk, H. M., Moore, R. A., and Morrison, J. M. 2002. "The design, implementation and use of Web-technologies to facilitate knowledge sharing: A 'real-world' application." In *Proceedings of the 2002 Command and Control Research and Technology Symposium,* Monterey, CA.

St. John, M., Manes, D. I., Oonk, H. M., and Ko, H. 1999. Workspace control diagrams and head-mounted displays as alternatives to multiple monitors in information-rich environments." In

Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting. Santa Monica, CA: Human Factors and Ergonomics Society. pp. 438-442.

St. John, M. Harris, W., and Osga, G. 1997. "Designing for multi-tasking environments: Multiple monitors vs. multiple windows." *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society. pp. 1313-1317.

Shrout, P. E., and Fleiss, J. L. 1979. "Intraclass correlations: Uses in assessing rate reliability." *Psychological Bulletin, 86*(2), pp. 420-428.

Sweller, J. 1988. "Cognitive load during problem solving: Effects on learning." *Cognitive Science*, 12, pp. 257-285.

Thackray, R. I., and Touchstone, R. M. 1991. "Effects of monitoring under high and low taskload on detection flashing and coloured radar targets." *Ergonomics, 34*, 1065-1081.

Whisenand, T. G., and Emurian, H. H. 1999. "Analysis of cursor movements with a mouse." *Computers in Human Behavior, 15*, 85-103.

APPENDIX A: USAGE/ACTIVITY SURVEY

In order to approximate realistic workload in the two experiments, a survey was sent to users in the Fleet (via an e-mail message) that asked questions about the amount of time they spent conducting various tasks. They were also asked to indicate any tasks that they thought should be included in a realistic warfighter task setting that were not asked about in the survey. Following a brief description of the upcoming LOE, the instructions to survey respondents were:

> To ensure that we provide a near-realistic setting, we would appreciate it if you would please provide some rough estimates of the following based on your experiences in Navy TFCC / CDC / CIC / War Room / etc. settings: IMPORTANT: Your responses should be independent of what you may have witnessed with regard to actual K-Desk use. We are after what tasks the average warfighter performs irrespective of the technology he or she has to deal with. Your inputs will help ensure that we provide a realistic task setting for the experiment.
>
> *Watchstanding Users*
>
> - How many chat windows are typically monitored at one time?
> - How much chat activity occurs (number of interactions per hour)?
> - How many high-priority e-mails are typically responded to (number per hour)?
> - How many tactical pictures are typically monitored? (multiple views or multiple systems)
> - How often does one refer to / monitor the tactical picture (per hour)?
> - How often does one answer / respond to information requests (per hour)? (from other watch standers, higher authority, subordinate commands)
> - How many information products are produced (per hour)? (PowerPoint briefs, Word reports / summaries, Excel charts, inputs to databases, naval messages, log entries, etc.)
> - How much time is spent reviewing documents / information / messages / the web (per hour)?
>
> *Non-watchstanding Users*
>
> - How many chat windows are typically monitored at one time?
> - How much chat activity occurs (number of interactions per hour)?
> - How many high-priority e-mails are typically responded to (number per hour)?
> - How many tactical pictures are typically monitored? (multiple views or multiple systems)
> - How often does one refer to / monitor the tactical picture (per hour)?
> - How often does one answer / respond to information requests (per hour)? (from other watch standers, higher authority, subordinate commands)
> - How many information products are produced (per hour)? (PowerPoint briefs, Word reports / summaries, Excel charts, inputs to databases, naval messages, log entries, etc.)
> - How much time is spent reviewing documents / information / messages / the web (per hour)?

RESULTS AND CORRESPONDENCE TO EXPERIMENTAL DESIGN

The results of the survey are shown in Figure A-1. No participant suggested any additional realistic tasks. The number of e-mails and chats sent, the total number of chat rooms monitored, and the

number of situation awareness questions asked (via chat plus e-mail) for each of the experiments was based on the results. The correspondence between real world activity and the experiment can also be seen in Figure A-1.

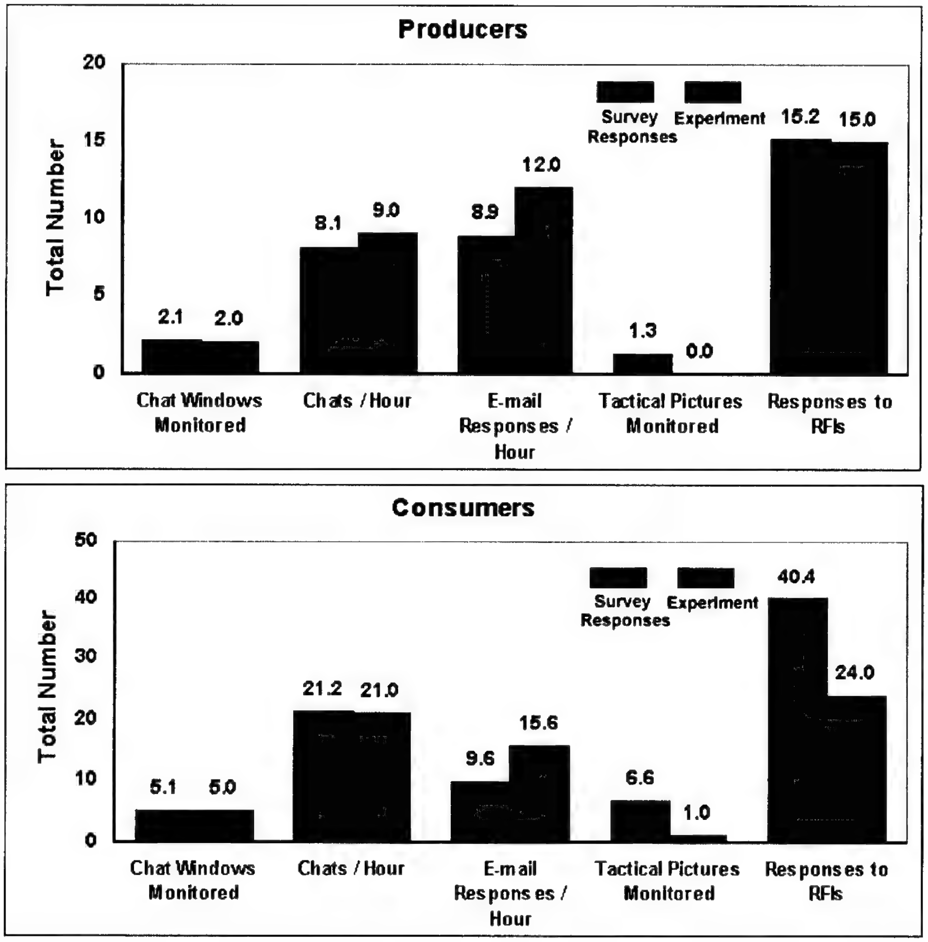

Figure A-1. The results of the Usage/Activity Survey and how well they corresponded with the design of the experiments.

APPENDIX B: RATER INSTRUCTIONS

Instructions for SME product evaluation

We conducted an experiment in which participants were asked to submit a product (MS Word document) to a fictional CJTF. The products were to contain information the participants deemed relevant and critical to the mission, based on information sources available to them. As a subject matter expert, you are being asked to evaluate these products on various dimensions.

Before you conduct product evaluations, it is important that you become *thoroughly* familiar with the task and materials that lead to the development of the information products. Please read the next two documents: *Experiment Instructions* and *Filling out the CJTF Product template.*

Below are the instructions that were read to all participants

Experiment Instructions

During this experiment, you will participate in a series of five 20-minute tasks based on five fictional missions. In each mission, your job will be to act as a Functional Component Commander, reporting to a CJTF. Your functional area of responsibility will change with each task, so in one task, you may be the MetOc FCC, but in the next task, you may be the Force Protect FCC.

During each 20-minute task, you will have access to a folder of documents and pictures relating to a specific mission in a specific part of the world. You'll be browsing through these documents to gain situation awareness. At the same time, you will also be monitoring two Chat rooms and one e-mail account. Messages that come through Chat and e-mail can be either new information or questions based on information available to you in the file folder. When you receive a question, please respond to it as quickly and accurately as possible.

Your task while monitoring the above information sources is to create a document to send to the CJTF that represents what you think is most important for the CJTF to be aware of, based on the material you are exposed to during the 20 minutes. You will be provided instructions and a template for creating this document later, as well as instructions on how to use Chat and e-mail.

Your performance on each task will be rated on the following:

1. The speed and accuracy of your Chat responses
2. The speed and accuracy of your e-mail responses
3. **The accuracy and quality of your product for the CJTF**

Additional Instructions that were read to the participants

FILLING OUT THE CJTF PRODUCT TEMPLATE

At the end of each 20-minute task, you will be submitting a product to the CJTF. This product is intended to represent what you think is most important for the CJTF to be aware of, based on the material you are exposed to during the 20 minutes. Consider that the CJTF has many tasks and little time; therefore, only include information that is relevant and important to the mission.

You will create your product using a very simple Word template that looks like this (show template). You may put as many or as few pieces of information as you feel is necessary by:

1. cutting and pasting from the files you have in your folder. You can cut and paste either entire documents or portions of a document (show example) Some documents are large and it is likely unnecessary to include the entire document for the purpose of briefing the CJTF;
2. cutting and pasting from chat or e-mail;
3. typing text directly into the document.

If you run out of table cells, you may add more space by using your Tab key. Keep in mind, however, that you only want to include the information that is most critical to share with the CJTF.

At the end of the 20-minute task, we will have you e-mail your template.

---- Sample Template ----

Bangladesh -- Logistics

[illegible]

Instructions for SME product evaluation (cont.)

Although each individual participated in five scenarios, you will only be evaluating information products from one of those scenarios.

As noted in the *Experiment Instructions*, the participant products are to be evaluated on accuracy and quality. We have operationalized those terms for the purposes of this study as follows:

Accuracy will be measured by:

- the number of relevant items included, and
- the number of irrelevant items that were included.

The above evaluations consider each part of an information product individually, without regard to format or appearance of the product.

Quality will be measured by:

- The degree to which the participant included the right information
- The degree to which the participant included the wrong information
- The degree to which the participant processed the information included

The above evaluations consider the information product as a whole, *without regard to format or appearance of the product.*

You will be provided with more detailed instructions on how to complete your evaluations. Before you begin, please read the *Evaluation Guidelines* on the following page.

Evaluation Guidelines

The following guidelines will help maintain the integrity and accuracy of your evaluations:

1. Do not discuss the evaluation of any products with anyone else, including other raters working on this project. This is to increase the reliability of your evaluations. If you have questions or concerns, please contact the investigators directly (see cover sheet). We are generally available to help you M-F 8:30 – 5:30 PST.

2. Allow enough time to complete each evaluation without major interruptions. If you are interrupted, make sure you review the evaluation criteria before continuing with your evaluation. Allow approximately 2 hours for completing both evaluations.

3. Review the evaluation response options each time you evaluate an item or product. This will help make sure you are using the same criteria each time you make a decision.

If you make a mistake, please draw a single line through the incorrect mark.

4. **Before beginning your evaluation, spend 45 minutes reviewing the content each participant was exposed to during the scenario. This includes a folder of documents and incoming communications (e-mail and MS Chat).**
    a. Some of the .html maps contained in the scenario folders have embedded links. If you see a colored line, circle or icon on a map, try clicking on it for more detailed information.
    b. Any information found in the incoming communications (chats and e-mails) file should be thought of as overriding the information found in the document folder.

Timeline for Evaluation Process (in Minutes)

Review Instructions:	15-20
Review Scenario Material (scenario 1):	45
Individual Item Evaluation (scenario 1):	30-45
Whole product Evaluation (scenario 1):	60-90
Complete Subject Matter Expert Info Sheet:	10

Total Time: 3 – 4 Hours (approx)

Instructions for Individual Item Evaluation

Once you have spent 45 minutes reviewing the content of a scenario folder, begin your Individual Item Evaluation. This is an evaluation of the relevance and accuracy of the content in each item, independent of other items. All items from all participants' information products have been compiled into one document. Items (in no particular order) are separated by black borders. Read each item, then mark an "x" in the **I**, **I/R** or **R** column to indicate your evaluation of the item relevance, using the following scale:

I	I/R	R
All or most of the information is "Irrelevant"	**Relevant information is accompanied by a significant amount of irrelevant information**	**All or most of the information is "Relevant"**

Examples:

- A map is presented, but there are no identifying labels or text that make it usable to the CJTF. **This item should be marked "I".**
- The same map is presented, but there are identifying labels or text, most of which is usable to the CJTF. **This item should be marked "R".**

- A source document contains information about the country of interest that may be important for the CJTF to have. Instead of pulling out the most key pieces, or paraphrasing, a large part of the document was copied into the item, including extraneous text. **This item should be marked "I/R".**
- A different source document contains information about the country of interest, but mostly nothing the CJTF needs. A portion of this document is copied into the item. **This item should be marked "I".**

- Weather information copied out of a weather log includes only the time-frame of interest. **This item should be marked "R".**
- Weather information copied out of a weather log includes several days in addition to the time-frame of interest. **This item should be marked "I/R".**

When evaluating:

- Items deemed inaccurate should be marked "I". This would include items that although relevant or accurate early in the scenario, became irrelevant or inaccurate later in the scenario (based on incoming Chats and E-mails).
  - Example: In a document -- "Plane X will fly" becomes inaccurate (and therefore irrelevant) when a Chat/e-mail states: "Plane X is under repair."
- Do not consider format or size of text or graphics.
- If unable to read a map/chart, refer to the original document in the scenario folder.
- Black borders signify boundaries between items, but, some items extend beyond one page. Use your judgment to determine if an item continues to another page (e.g., sentence ends abruptly on one page, then continues on the next).
- Consider all content contained in a single box to be part of one item.
- **Several items may appear to be identical, but most likely have slightly different content. Please consider these items individually.**

Instructions for Whole Product Evaluation

Once you have completed the Individual Item Evaluation, you can begin your Whole Product Evaluation. This is an evaluation of the overall quality of each participants' product. Read the entire product, then mark an "x" in either the appropriate column to indicate your evaluation of the product quality, based on each of the following three scales:

Including the right information

1	2	3
Includes none (or little) of the important / critical information that should have been provided to the CJTF.	**Includes a fair amount of the important / critical information that should have been provided to the CJTF.**	**Includes most (or all) of the important / critical information that should have been provided to the CJTF.**

Including the wrong information

1	2	3
Includes mostly (or only) information that was not important / critical to provide to the CJTF.	**Includes some information that was not important / critical to provide to the CJTF.**	**Includes no (or little) information that was not important / critical to provide to the CJTF.**

Processing the information

1	2	3
Does not process (interpret or paraphrase) information or make connections between items for the CJTF's understanding.	**Does some processing (interpreting or paraphrasing) of information and/ or makes connections between items for the CJTF's understanding.**	**Most or all of the content is processed (interpreted or paraphrased). Connections are made between information items for the CJTF's understanding.**

When evaluating:

- For each of the above scales, consider the product as a whole. Do not base the evaluation on just a few information items.
- Do not consider format or size of text or graphics.
- If unable to read a map or chart, refer to the original document in the scenario folder.
- Keep in mind that quality is more important than quantity.

APPENDIX C: RESULTS TABLES

The tables in this appendix correspond to the figures presented in the report text.

Table C- 1. Preferred number of monitors indicated by participants in the Producer experiment (corresponds to Figure 3).

	Number of Monitors							
	One	Two	Three	Three to Four	Four	Five	Six	Four, Five, or Six
Number of Participants	0	1	2	3	9	4	3	1

Table C- 2. Speed, accuracy, misses, and adjusted scores for e-mail inquiry responses for the different monitor conditions in the Producer experiment (corresponds to Figure 4).

	Number of Monitors				
	1	**2**	**3**	**4**	**6**
Response Times (minutes)	2.62	2.96	2.90	2.29	2.59
Proportion Correct	.72	.63	.62	.70	.60
Number of Missed Questions	.90	.73	1.10	1.00	.90
Adjusted Accuracy Scores	1.66	1.52	1.50	1.59	1.52

Table C- 3. Speed, accuracy, misses, and adjusted scores for chat inquiry responses for the different monitor conditions in the Producer experiment (corresponds to Figure 5).

	Number of Monitors				
	1	**2**	**3**	**4**	**6**
Response Times (minutes)	2.87	3.75	2.93	2.95	2.76
Proportion Correct	.56	.63	.50	.79	.67
Number of Missed Questions	.50	.54	.50	.25	.42
Adjusted Accuracy Scores	1.17	1.29	1.08	1.63	1.33

Table C- 4. The average composite product accuracy scores for the different monitor conditions in the Producer experiment (corresponds to Figure 6).

	Number of Monitors				
	1	**2**	**3**	**4**	**6**
Average Product Accuracy Scores	9.76	10.92	8.56	11.53	11.79

Table C- 5. The average whole product quality scores for the different monitor conditions in the Producer experiment (corresponds to Figure 7).

	Number of Monitors				
	1	2	3	4	6
Average Whole Product Right Scores	1.75	1.70	1.83	1.84	1.83
Average Whole Product Wrong Scores	2.25	2.35	2.38	2.25	2.42
Average Whole Product Processing Scores	1.39	1.53	1.42	1.31	1.35

Table C- 6. Preferred number of monitors indicated by participants in the Consumer experiment (corresponds to Figure 11).

	Number of Monitors						
	One	Two	Three	Three to Four	Four	Five	Six
Number of Participants	0	0	5	5	10	3	6

Table C- 7. Speed, accuracy, misses, and adjusted scores for e-mail inquiry responses for the different monitor conditions in the Consumer experiment (corresponds to Figure 12).

	Number of Monitors				
	1	2	3	4	6
Response Times (minutes)	2.83	2.58	2.46	2.14	2.30
Proportion Correct	.73	.70	.72	.74	.80
Number of Missed Questions	.07	.30	.27	.20	.20
Adjusted Accuracy Score	1.68	1.64	1.64	1.68	1.63

Table C- 8. Speed, accuracy, misses, and adjusted scores for chat inquiry responses for the different monitor conditions in the Consumer experiment (corresponds to Figure 13).

	Number of Monitors				
	1	2	3	4	6
Response Times (minutes)	2.59	2.40	2.46	2.19	2.27
Proportion Correct	.53	.58	.64	.66	.66
Number of Missed Questions	1.13	.93	.90	.73	.83
Adjusted Accuracy Score	1.14	1.27	1.34	1.39	1.38

Table C- 9. Mean number of web accesses during a block in the Consumer experiment (corresponds to Figure 14).

	Number of Monitors				
	1	2	3	4	6
Number of Web Accesses / Scenario	12.53	13.63	12.53	13.10	15.60

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-01-0188*

1. REPORT DATE *(DD-MM-YYYY)* 03–2003	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)*

4. TITLE AND SUBTITLE
KNOWLEDGE DESK LIMITED OBJECTIVE EXPERIMENT (LOE)

5a. CONTRACT NUMBER N66001-99-D-0050
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER 0602233N
5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

6. AUTHORS

H. M. Oonk, J. H. Schermerhorn, D. Glaser
Pacific Science and Engineering Group, Inc.

J. G. Morrison
SSC San Diego

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Pacific Science and Engineering Group, Inc.
6310 Greenwich Drive, Suite 200
San Diego, CA 92122

SSC San Diego
San Diego, CA 92152–5001

8. PERFORMING ORGANIZATION REPORT NUMBER
TR 1896

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217–5660

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
This is the work of the United States Government and therefore is not copyrighted. This work may be copied and disseminated without restriction. Many SSC San Diego public release documents are available in electronic format at http://www.spawar.navy.mil/sti/publications/pubs/index.html

14. ABSTRACT

This report was prepared as part of SSC San Diego's ongoing Command 21 project. One goal of this project is the development and operational evaluation of multi-headed displays known as the Knowledge Desks for use in advanced military command centers ashore and at sea. This report describes an evaluation in the continuing efforts directed toward this goal.

15. SUBJECT TERMS
Mission Area: Command and Control
Knowledge desks
Command centers

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	52	Wayne L. Patterson
					19B. TELEPHONE NUMBER *(Include area code)* (619) 553–1423

INITIAL DISTRIBUTION

20012	Patent Counsel	(1)
202753	Archive/Stock	(4)
202752	Library	(2)
2027	M. E. Cathcart	(1)
202753	D. R. Richter	(1)
240	R. Smith	(1)
2403	J. Iaia	(1)
2405	S. Hearold	(1)
2411	L. Duffy	(1)
242	R. Jaffee	(1)
244	J. Martin	(1)
2441	R. Fleming	(1)
2441	J. Grossman	(1)
2441	J. Morrison	(10)
2441	T. Tiernan	(1)
244210	R. Smillie	(1)
244215	J. Kaiwi	(1)
272	J. Audia	(1)

Defense Technical Information Center
Fort Belvoir, VA 22060–6218 (4)

SSC San Diego Liaison Office
Arlington, VA 22202–4804

Center for Naval Analyses
Alexandria, VA 22302–0268

Office of Naval Research
ATTN: NARDIC (Code 362)
Arlington, VA 22217-5660

Government-Industry Data Exchange
Program Operations Center
Corona, CA 91718–8000

Office of Naval Research
Arlington, VA 22217–5000 (2)

Office of Naval Research
Arlington, VA 22217–5660 (4)

Naval War College
Newport, RI 02841-1207 (6)

Bureau of Naval Personnel
Washington, DC 20370-0100

Naval Air Warfare Center
Training Systems Division
Orlando, FL 32826–3275 (2)

Commander Carrier Group ONE
San Diego CA 92135 (2)

Commander in Chief, U.S. Pacific Fleet
Pearl Harbor, HI 96860-3131 (2)

Commander in Chief, U.S. Atlantic Fleet
Norfolk, VA 23551-2487

Space and Naval Warfare Systems Center
North Charleston, SC 29419-9022

Space and Naval Warfare Systems Command
San Diego, CA 92110-3127 (3)

Navy Warfare Development Command
Newport, RI 02841-1207

Office of the Secretary of Defense
Executive Support Center
Washington, D.C. 20330 (3)

Naval Postgraduate School
Monterey, CA 93943

C4I Academic Group
Monterey, CA 93943

Aptima, Inc.
Woburn, MA 01801 (2)

Northrop Grumman IT/TASC
Chantilly, VA 20151

Northrup Grumman Corporation
San Diego, CA 92110-5151

Director/Human Factors Division
Downsview, Ontario M3M 3B9

Centre De Recherches Pour La Defense,
Valcartier
Defence Research Establishment Valcartier
Courcelette, QUE. Canada, GOA 1RO

Arizona State University
Department of Psychology
Tempe, AZ 85287-1104

Klein Associates Inc.
Fairborn, Ohio 45324-3987

San Diego State University
San Diego, CA 92120

Pacific Science and Engineering
San Diego, CA 92122

Evidence Based Research, Inc.
Vienna, VA 22182-2216

Science Applications International Corporation
Arlington, VA 22203

Sonalysts, Inc.
Waterford, CT 06385